PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification: **C07K 14/705, A61K 38/17, A61P 25/00, C07K 14/47, G06F 17/50** | **A1**

(11) International Publication Number: **WO 00/37500**

(43) International Publication Date: 29 June 2000 (29.06.2000)

(21) International Application Number: PCT/CA99/01209

(22) International Filing Date: 17 December1999 (17.12.1999)

(30) Priority Data:
60/112,929 18 December1998 (18.12.1998) US

(60) Parent Application or Grant
MOUNT SINAI HOSPITAL [/]; (). STAPLETON, David [/]; (). SICHERI, Frank [/]; (). STAPLETON, David [/]; (). SICHERI, Frank [/]; (). TANDAN, Susan, I. ; ().

Published

(54) Title: THREE DIMENSIONAL STRUCTURE OF A STERILE ALPHA MOTIF DOMAIN

(54) Titre: STRUCTURE TRIDIMENSIONNELLE D'UN DOMAINE DE MOTIF ALPHA STERILE

(57) Abstract

The invention relates to the three dimensional structure of sterile alpha motif (Sam) domain. The atomic coordinates that define the structure and any compounds bound to the structure enable the determination of the three dimensional structures of SAM domains with unknown structure, and the identification of modulators of a SAM domain.

(57) Abrégé

L'invention concerne la structure tridimensionnelle d'un domaine de motif alpha stérile (maS). Les coordonnées atomiques définissant la structure et les éventuels composés liés à la structure permettent la détermination des structures tridimensionnelles des domaines MAS de structure inconnue, et l'identification de modulateurs d'un domaine MAS.

PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification [7]:
C07K 14/705, 14/47, G06F 17/50, A61K 38/17, A61P 25/00

A1

(11) International Publication Number: WO 00/37500

(43) International Publication Date: 29 June 2000 (29.06.00)

(21) International Application Number: PCT/CA99/01209

(22) International Filing Date: 17 December 1999 (17.12.99)

(30) Priority Data:
60/112,929 18 December 1998 (18.12.98) US

(71) Applicant *(for all designated States except US)*: MOUNT SINAI HOSPITAL [CA/CA]; Samuel Lunenfeld Research Institute, Office of Technology Transfer & Industrial Liaison, 600 University Avenue, Toronto, Ontario M5G 1X5 (CA).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: STAPLETON, David [AU/AU]; St. Vincent's Institute of Medical Research, 9 Princes Street, Fitzroy, Victoria 3065 (AU). SICHERI, Frank [CA/CA]; Apartment 2110, 77 Gerrard Street West, Toronto, Ontario M5G 2A1 (CA).

(74) Agents: TANDAN, Susan, I. et al.; Swabey Ogilvy Renault, Suite 1407, 77 Bloor Street West, Toronto, Ontario M5S 1M2 (CA).

(81) Designated States: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, DE, DK, DM, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published
With international search report.
Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: THREE DIMENSIONAL STRUCTURE OF A STERILE ALPHA MOTIF DOMAIN

(57) Abstract

The invention relates to the three dimensional structure of sterile alpha motif (Sam) domain. The atomic coordinates that define the structure and any compounds bound to the structure enable the determination of the three dimensional structures of SAM domains with unknown structure, and the identification of modulators of a SAM domain.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav	TM	Turkmenistan
BF	Burkina Faso	GR	Greece		Republic of Macedonia	TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's	NZ	New Zealand		
CM	Cameroon		Republic of Korea	PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

Description

TITLE: Three Dimensional Structure of a Sterile Alpha Motif Domain

FIELD OF THE INVENTION

The invention relates to the three dimensional structure of a sterile alpha motif (Sam) domain. The atomic coordinates that define the structure and any compounds bound to the structure enable the determination of homologues, the three dimensional structures of polypeptides with unknown structure, and the identification of modulators of a SAM domain.

BACKGROUND OF THE INVENTION

The Eph family of receptor tyrosine kinases have been implicated in the control of axon guidance [Henkemeyer, 1996; Orioli, 1996], cell migration [Krull, 1997], patterning of the nervous system [Xu, 1996] and angiogenesis [Wang, 1998], and are activated by clustering into dimers or tetramers [Stein, 1998]. However, the cell-surface ligands for Eph receptors (ephrins) apparently lack an intrinsic ability to induce receptor oligomerization [Lackmann, 1997]. Factors that influence receptor aggregation include the pre-clustering of ephrins [Davis, 1994], the homotypic interaction between the extracellular domains of two receptor chains [Lackmann, 1998], and the binding of PDZ domain containing proteins to the receptor's C-terminus [Hock, 1998].

All Eph receptors have a Sterile Alpha Motif (SAM) domain within their cytoplasmic regions. The SAM domain was identified as a conserved sequence present in a small set of yeast sexual differentiation proteins referred to as the Sterile Alpha Mating factors [Ponting, 1995; Schultz, 1997]. In ETS family transcription factors this sequence has also been termed the Pointed domain [Klambt, 1993]. The domain is found in a variety of proteins, many of which contain catalytic domains or recognized protein interaction domains. SAM domains are almost always located at a protein's N- or C-terminus. A highly conserved SAM domain is located in the cytoplasmic region of Eph receptors (approx. 50 % identity over 14 family members), C-terminal to the catalytic domain and followed by only 5 residues that form a potential PDZ domain binding site [Hock, 1998]. Amongst receptor tyrosine kinases, the presence of a cytoplasmic module other than the protein kinase domain is unique to Eph receptors.

The SAM domain can function as a protein interaction module through an ability to homo- and hetero-dimerize with other SAM domains [Jousset, 1997; Peterson, 1997; Tu, 1997; Kyba, 1998]. This dimerizing property elicits oncogenic activation of chimeric proteins arising from translocation of the SAM domain of TEL to coding regions of the βPDGF receptor [Golub, 1994); Abl [Golub, 1996], and JAK2 protein kinases [Lacronique, 1997] or the AML1 transcription factor [Golub, 1995]. A functional role in mediating homo and hetero-typic dimerization has been shown for SAM domains in the transcription factor TEL [Jousset, 1997], members of the polycomb group of transcriptional repressors (RAE28, Scm and ph) [Peterson, 1997], the protein kinase Byr2p [Tu, 1997], and the α and β isoforms of the liprin scaffolding proteins [Serra-Pages, 1998].

SUMMARY OF THE INVENTION

Broadly stated, the present invention relates to the three-dimensional structure of one or more SAM domains. The three-dimensional structures may be complexed with one or more compounds. The defined boundaries and properties of the structures and any of the compounds bound

to it are pertinent to methods for determining the three-dimensional structures of polypeptides with unknown structure, and to methods that identify modulators of SAM domain function. These modulators are potentially useful as therapeutics for diseases, including (but not limited to) cell proliferative diseases, such as cancer, angiogenesis, atherosclerosis, and arthritis, and diseases associated with the nervous system.

Broadly stated the present invention relates to a crystalline form of a polypeptide corresponding to one or more SAM domains, preferably one or more SAM domains of an Eph receptor, preferably of EphA. The crystalline form may comprise one or more heavy metal atoms, or at least one compound. In a preferred embodiment, a unit cell of the crystalline form of the invention has dimensions of about a=b= 77.14 ± .03 angstroms, c= 24.3 ± .04 angstroms.

The invention also relates to a method of forming a crystalline form of the invention comprising

(a) mixing a volume of a SAM domain with a reservoir solution; and

(b) incubating the mixture obtained in step (a) over the reservoir solution in a closed container under conditions suitable for crystallization.

The invention also features a method of determining three dimensional structures of polypeptides with unknown structure comprising the step of applying the structural atomic coordinates of a crystalline form of one or more SAM domains of the invention.

Methods are also provided for identifying a potential modulator of a SAM domain function preferably a SAM domain of an Eph receptor function by docking a computer representation of a structure of a compound with a computer representation of a structure of one or more SAM domains of the invention preferably a SAM domain of an Eph receptor that is defined by the atomic structural coordinates described herein. In an embodiment the method comprises the following steps:

(a) docking a computer representation of a compound from a computer data base with a computer representation of a selected site on a SAM domain, preferably a SAM domain of an Eph receptor,to obtain a complex;

(b) determining a conformation of the complex with a favourable geometric fit and favourable complementary interactions; and

(c) identifying compounds that best fit the selected site as potential modulators of SAM domain function.

In another embodiment the method comprises the following steps:

(a) modifying a computer representation of a compound complexed with a selected site on a SAM domain, preferably a SAM domain of an Eph receptor, by deleting or adding a chemical group or groups;

(b) determining a conformation of the complex with a favourable geometric fit and favourable complementary interactions; and

(c) identifying a compound that best fits the selected site as a potential modulator of a SAM domain.

In still another embodiment the method comprises the following steps:

(a) selecting a computer representation of a compound complexed with a selected site on a SAM domain, preferably a SAM domain of an Eph receptor; and

(b) searching for molecules in a data base that are similar to the compound using a searching computer program, or replacing portions of the compound with similar chemical structures from a data base using a compound building computer program.

The invention also features a potential modulator of a function of a SAM domain preferably a SAM domain of an Eph receptor identified by the methods of the invention, and a method of treating a disease associated with a SAM domain preferably a SAM domain of an Eph receptor with inappropriate activity in a cellular organism, comprising:

(a) administering a modulator identified using the methods of the invention in an acceptable pharmaceutical preparation; and

(b) activating or inhibiting a SAM domain function to treat the disease.

The invention also provides peptides that mediate SAM domain function.

BRIEF DESCRIPTION OF THE DRAWINGS

The invention will now be described in relation to the drawings in which:

Figure 1A shows a sequence alignment of SAM domains from selected proteins (SEQ. ID. NOS. 1 to 21);

Figure 1B shows a selection of multi-domain proteins containing SAM domain (S);

Figure 2A is a ribbons depiction of the SAM homo-dimer viewed down the twofold symmetry axis;

Figure 2B is a ribbons depiction of the SAM homo-dimer viewed perpendicular to the symmetry axis;

Figure 2C is a ribbons stereo view highlighting the dimer interface region;

Figure 3A is a molecular surface and worm representation of the SAM homodimer;

Figure 3B is a molecular surface and worm representation of the SAM homodimer; and

Figure 4 is a gel filtration elution profile of wild type and single or double site mutants of the EphA4 receptor SAM domain.

DETAILED DESCRIPTION OF THE INVENTION

DEFINITIONS:

Unless otherwise indicated, all terms used herein have the same meaning as they would to one skilled in the art of the present invention. Practitioners are particularly directed to Current Protocols in Molecular Biology (Ansubel) for definitions and terms of the art.

Abbreviations for amino acid residues are the standard 3-letter and/or 1-letter codes used in the art to refer to one of the 20 common L-amino acids. Likewise abbreviations for nucleic acids are the standard codes used in the art.

The term "crystalline form" in the context of the invention, is a crystal formed from an aqueous solution comprising a purified polypeptide comprising one or more SAM domains, preferably a SAM domain of an Eph receptor. A crystalline form of a SAM domain is characterized

as being capable of diffracting x-rays in a pattern defined by one of the crystal forms depicted in Blundel et al 1976, Protein Crystallography, Academic Press. A crystalline form may include a crystal structure in association with one or more heavy-metal atoms i.e. a derivative crystal, or a crystal structure in association with one or more compounds i.e. a co-crystal.

The term "association" refers to a condition of proximity between a chemical entity or compound or portions or fragments thereof, and a SAM domain or portions or fragments thereof. The association may be non-covalent i.e. where the juxtaposition is energetically favored by for example, hydrogen-bonding, van der Waals, or electrostatic or hydrophobic ineractions, or it may be covalent.

The term "heavy-metal atoms" refers to an atom that is a transition element, a lanthanide metal, or an actinide metal. Lanthanide metals include elements with atomic numbers between 57 and 71, inclusive. Actinide metals include elements with atomic numbers between 89 and 103, inclusive.

The term "Eph receptor" refers to a subfamily of closely related transmembrane receptor tyrosine kinases related to Eph, a receptor named for its expression in an erythropoietin-producing human hepatocellular carcinomas cell line. The receptors contain cell adhesion-like domains on their extracellular surface. The Eph subfamily receptor tyrosine kinases are more specifically characterised as encoding a structurally related cysteine rich extracellular domain containing a single immunoglobulin (Ig)-like loop near the N-terminus and two fibronectin III (FN III) repeats adjacent to the plasma membrane. The Eph receptors are divided into two groups based on the relatedness of their extracellular domain sequences. The grouping also corresponds to the ability of the receptors to bind preferentially to the ephrin-A or ephrin-B proteins. The group that includes receptors interacting preferentially with ephrin A proteins is called EphA and includes EphA1 (also known as Eph and Esk), EphA2 (also known as Eck, Myk2, Sek2), EphA3 (also known as Cek4, Mek4, Hek, Tyro4, Hek4), EphA4 (also known as Sek, Sek1, Cek8, Hek8, Tyro1), EphA5 (also known as Ehk1, Bsk, Cek7, Hek7, and Rek7), EphA6 (Ehk2, and Hek12) EphA7 (also known as Mdk1, Hek11, Ehk3, Ebk, Cek11), and EphA8 (also known as Eek, Hek3). The group that includes receptors interacting preferentially with ephrin B proteins is called Eph B and includes EphB1 (also known as Elk, Cek6, Net, Hek6), EphB2 (also known as Cek5, Nuk, Erk, Qek5, Tyro5, Sek3, hek5, Drt), EphB3 (also known as Cek10, Hek2, Mdk5, Tyro6, and Sek4), EphB4 (also known as Htk, Myk1, Tyro11, Mdk2), EphB5 (also known as Cek9, Hek9), and EphB6 (also known as Mep).

"Ephrin" refers to a class of ligands which are anchored to the cell membrane through a transmembrane domain, and bind to the extracellular domain of an Eph receptor, facilitating dimerization and autophosphorylation of the receptor and autophosphorylation of the ligand. The ephrins which are targeted in the methods of the invention are those that bind to and activate (i.e. phosphorylate) an EphA or an EphB receptor, preferably an EphA receptor. The ephrin-A ligands (GPI-anchored ligands) are ephrin-A (also known as B61, LERK1, EFL-1), ephrin-A2 (also known as LERK6, Elf1, mCek7-L, cElf1), ephrin-A3 (also known as LERK3, Ehk1-L, and EFL-2), ephrin-A4 (also known as LERK4, EFL-4, mLERK4), ephrin-A5 (AL1, LERK7, EFL-5, mAL1, [rLERK7], RAGS), and the ephrin-B ligands (transmembrane ligands) are ephrin-B1 (also known as LEKR2,

ELK-L, EFL-3, Cek5-L, Stra1, [LERK2]), ephrin-B2 (also known as LERK5, HTK-L, NLERK1, Elf2, Htk-L), and ephrin-B3 (also known as LERK8, ELK-L3, NLERK2, EFL-6, Elf3, [rELK-L3]).

The term "SAM domain" refers to a region known as the Sterile Alpha Motif (SAM) domain within the cytoplasmic regions of all Eph receptors (Figure 1B), and in other proteins such as TEL [Jousset, 1997], members of the polycomb group of transcriptional repressors (RAE28, Scm and ph) [Peterson, 1997], the protein kinase Byr2p [Tu, 1997], the α and β isoforms of the liprin scaffolding proteins [Serra-Pages, 1998], and tankyrase (Smith, S. et al, Science 282: 1484-1487, 1998, Acession AF082556). The SAM domain was identified as a conserved sequence present in a small set of yeast sexual differentiation proteins referred to as the Sterile Alpha Mating factors [Ponting, 1995; Schultz, 1997]. In ETS family transcription factors this sequence has also been termed the Pointed domain [Klambt, 1993]. Extensive database searching and sequence alignment analysis (Figure 1A) reveals that this domain is found in a variety of proteins, many of which contain catalytic domains or recognized protein interaction domains (Figure 1B). SAM domains are almost always located at a protein's N- or C-terminus. A highly conserved SAM domain is located in the cytoplasmic region of Eph receptors (approximately 50 % identity over 14 family members), C-terminal to the catalytic domain and followed by only 5 residues that form a potential PDZ domain binding site [Hock, 1998]. The term also includes amino acid sequences having substantial sequence identity to a SAM domain, a mutant, or a subunit of a SAM domain. Preferably the SAM domain is an "Eph SAM domain" i.e. a SAM domain of an Eph receptor.

"SAM domain structure" or "SAM domain three dimensional structure" refers to the three dimensional structure of a purified polypeptide comprising one or more SAM domains, preferably a crystalline form.

As applied to polypeptides, the term " substantial sequence identity" means that two peptide sequences, when optimally aligned, such as by the programs GAP or BESTFIT using default gap, share at least 80 percent sequence identity, preferably at least 90 percent sequence identity, more preferably at least 95 percent sequence identity or more. Preferably, residue positions which are not identical differ by conservative amino acid substitutions. For example, the substitution of amino acids having similar chemical properties such as charge or polarity are not likely to effect the properties of a protein. Examples include glutamine for asparagine or glutamic acid for aspartic acid.

The term "mutant" refers to a polypeptide that is obtained by replacing at least one amino acid residue in a native SAM domain with a different amino acid residue. Mutation can be accomplished by adding and/or deleting amino acid residues within the native SAM domain. A mutant may or may not be functional.

The term "function" refers to the ability of a modulator to enhance or inhibit the association between a SAM domain and a compound.

The term "atomic structural coordinates" as used herein refers to a data set that defines the three dimensional structure of a molecule or molecules (e.g. unit cell axial lengths, space group). Structural coordinates can be slightly modified and still render nearly identical three dimensional structures. A measure of a unique set of structural coordinates is the root-mean-square deviation of

the resulting structure. Structural coordinates that render three dimensional structures that deviate from one another by a root-mean-square deviation of less than 1.5 Å may be viewed by a person of ordinary skill in the art as identical. Structural coordinates for a SAM domain are in Table 2.

The term "unit cell" refers to the smallest and simplest volume element (i.e. parallelpiped-shaped block) of a crystal that is completely representative of the unit of pattern of the crystal. The unit cell axial lengths are represented by a, b, and c where a = x axis, b = y axis, and c = z axis. Those of skill in the art understand that a set of atomic coordinates determined by X-ray crystallography is not without standard error.

The term "space group" refers to the symmetry of a unit cell. In a space group designation the capital letter indicates the lattice type and the other symbols represent symmetry operations that can be carried out on the unit cell without changing its appearance.

The term "purified" in reference to a polypeptide, does not require absolute purity such as a homogenous preparation rather it represents an indication that the sequence is relatively purer than in the natural environment. Generally, a purified polypeptide is substantially free of other proteins, lipids, carbohydrates, or other materials with which it is naturally associated, preferably at a functionally significant level for example at least 85% pure, more preferably at least 95% pure, most preferably at least 99% pure. A skilled artisan can purify a polypeptide comprising a SAM domain using standard techniques for protein purification. A substantially pure polypeptide comprising a Sam domain will yield a single major band on a non-reducing polyacrylamide gel. The purity of the SAM domain polypeptide can also be determined by amino-terminal amino acid sequence analysis.

Three Dimensional Structure of SAM Domain

The present invention provides a purified SAM domain three dimensional structure. In an embodiment the structure is a crystalline form. A SAM domain structure may comprise one or more SAM domains in a unit cell, preferably two, three or four SAM domains. In a preferred embodiment, a SAM domain is arranged in a crystallline manner in a space group $P6_4$ so as to form a unit cell of dimensions a=b= 77.14 angstroms, c= 24.37 angstroms and which effectively diffracts X-rays for determination of the atomic coordinates of the SAM domain to a resolution of about 2.9 angstroms. The 3-dimensional structure of a preferred SAM domain of the invention is shown in Figures 2 and 3.

A crystalline form includes native crystals, derivative crystals, and co-crystals. The native crystals generally comprise substantially pure polypeptides comprising one or more SAM domains in crystalline form. It is understood that the crystalline form is not limited to naturally occurring or native SAM domains but includes mutants of native SAM domains obtained by replacing at least one amino acid residue in a native SAM domain with a different amino acid residue or by adding or deleting amino acid residues within the native polypeptide, and having substantially the same three dimensional structure as the native SAM domain from which the mutant is derived i.e. having a set of atomic structural coordinates that have a root mean square deviation of less than or equal to about 2Å when superimposed with the atomic structure coordinates of the native SAM domain from which the mutant is derived when at least 50% to 100% of the atoms of the native SAM domain are included in

the superimposition. It should be noted that the mutants contemplated herein need not exhibit SAM domain activity.

The derivative crystals of the invention generally comprise a crystalline SAM domain in covalent association with one or more heavy metal atoms. The SAM domain may correspond to a native or mutated SAM domain. Heavy metal atoms useful for providing derivative crystals include by way of example, and not limitation gold, mercury, etc.

The invention features a crystalline form of a SAM domain in association with one or more compounds. The association may be covalent or non-covalent. These types of crystalline forms are referred to herein as co-crystals. The compound may be any organic molecule, and it may modulate the function of a SAM domain by for example inhibiting or enhancing its function, or it may be an analogue of a SAM domain. It is preferred that the geometry of the compound and the interactions formed between the compound and the SAM domain provide high affinity binding between the two molecules. High affinity binding is preferably governed by a dissociation equilibrium constant on the order of 10^{-6} or less.

Method for Preparing Crystal Forms of SAM Domain

The invention also features a method for creating the crystalline SAM domain structures described herein. The method may utilize a polypeptide comprising a SAM domain described herein to form a crystal. A polypeptide used in the method may be chemically synthesized in whole or in part using techniques that are well-known in the art. Alternatively, methods are well known to the skilled artisan to construct expression vectors containing the native or mutated SAM domain coding sequence and appropriate transcriptional/translational control signals. These methods include *in vitro* recombinant DNA techniques, synthetic techniques, and *in vivo* recombination/genetic recombination. See for example the techniques described in Sambrook et al. (Molecular Cloning: A Laboratory Manual, 2nd Edition, Cold Spring Harbor Laboratory press (1989)), and other laboratory textbooks.

Crystals are grown from an aqueous solution containing the purified and concentrated SAM domain polypeptide by a variety of conventional processes. These processes include batch, liquid, bridge, dialysis, vapor diffusion, and hanging drop methods. (See for example, McPherson, 1982 John Wiley, New York; McPherson, 1990, Eur. J. Biochem. 189: 1-23; Webber. 1991, Adv. Protein Chem. 41:1-36). Generally, the native crystals of the invention are grown by adding precipitants to the concentrated solution of the SAM domain polypeptide. The precipitants are added at a concentration just below that necessary to precipitate the protein. Water is removed by controlled evaporation to produce precipitating conditions, which are maintained until crystal growth ceases.

In an embodiment of the invention, the method generally comprises the steps of

(a) mixing a volume of polypeptide solution with a reservoir solution; and

(b) incubating the mixture obtained in step (a) over the reservoir solution in a closed container, under conditions suitable for crystallization.

For crystals of the invention, it has been found that hanging drops containing about 1µl of SAM domain polypeptide (50-150 mg/ml, preferably 100 mg/ml, in 5-2-mM, preferably 7mM Hepes

pH 5.5 to 9, preferably 7.5) and equal volumes of reservoir buffer (50-150 mM, preferably 100mM cacodylate pH 5.5 to 7.5, preferably 6.5; 5-10% preferably 7% (w/v) PEG 8000; and 10-30%, preferably 20% (v/v) ethylene glycol) suspended overnight at room temperature provide crystals suitable for high resolution X-ray structure determination. It will be appreciated that the above-described crystallization conditions can be varied and such variations can be used alone or in combination. For example other buffer solutions such as Tris-HCL buffer may be used.

Derivative crystals of the invention can be obtained by soaking native crystals in a solution containing salts of heavy metal atoms. Co-crystals of the invention can be obtained by soaking a native crystal in a solution containing a compound that binds the SAM domain, or they can be obtained by co-crystallizing the SAM domain polypeptide in the presence of one or more compounds that bind to the SAM domain.

Once the crystal is grown it can be placed in a glass capillary tube and mounted onto a holding device connected to an X-ray generator and an X-ray detection device. Collection of X-ray diffraction patterns are well documented by those skilled in the art (See for example, Ducruix and Geige, 1992, IRL Press, Oxford, England). A beam of X-rays enter the crystal and diffract from the crystal. An X-ray detection device can be utilized to record the diffraction patterns emanating from the crystal. Suitable devices include the Marr 345 imaging plate detector system with an RU200 rotating anode generator.

Methods for obtaining the three dimensional structure of the crystalline form of a molecule or complex are described herein and known to those skilled in the art (see Ducruix and Geige, supra). Generally, the unit cell dimensions and orientation in the crystal can be determined from the spacing between the diffraction emissions as well as the patterns made from the emissions. The symmetry of the unit cell in the crystal is also determined. Each diffraction pattern emission is characterized as a vector and the data collected at this stage determines the amplitude of each vector. The phases of the vectors may be determined by the isomorphous replacement method where heavy atoms soaked into the crystal are used as reference points in the X-ray analysis (see for example, Otwinowski, 1991, Daresbury, United Kingdom, 80-86). The phases of the vectors may also be determined by molecular replacement (see for example, Naraza, 1994, Proteins 11:281-296). The amplitudes and phases of vectors from the crystalline form of an Eph SAM domain, preferably an EphA4 SAM domain, determined in accordance with these methods can be used to analyze other crystalline SAM domains.

The unit cell dimensions and symmetry, and vector amplitude and phase information can be used in a Fourier transform function to calculate the electron density in the unit cell i.e. to generate an experimental electron density map. This may be accomplished used the PHASES package (Furey, 1990). Amino acid sequence structures are fit to the experimental electron density map (ie. model building) using computer programs (e.g. Jones, TA. et al, Acta Crystallogr A47, 100-119, 1991) to calculate a theoretical electron density map. The theoretical and experimental electron density maps can be compared and the agreement between the maps can be described by a parameter referred to as R-factor. A high degree of overlap in the maps is represented by a low value R-factor. The R-factor can be minimized by using computer programs that refine the theoretical electron density map. For

example, the XPLOR program, developed by Brunger (1992, Nature 355:472-475) can be used for model refinement.

A three dimensional structure of the molecule or complex may be described by atoms that fit the theoretical electron density characterized by a minimum R value. Files can be created for the structure that defines each atom by coordinates in three dimensions.

Identification of Homologues

The knowledge of the three dimensional structure of a SAM domain, in particular the EphA4 SAM domain, enables one skilled in the art to identify homologues. This is achieved by searches of three-dimensional databases. Since structural folds are conserved to a greater extent than sequence, one may identify homologues with very little sequence similarity. Programs that provide this type of database searching are known in the art and include Dali. The structural coordinates of a protein structure are submitted and the program performs a multiple structural alignment with proteins in the protein data bank.

Methods for Determining Three Dimensional Structures

The structure coordinates of a SAM domain structure described herein can be used as a model for determining the three dimensional structures of additional native or mutated SAM domains with unknown structure, as well as the structures of co-crystals of SAM domains with compounds such as modulators (e.g. agonists or antagonists). The structure coordinates and models of a SAM domain three dimensional structure can also be used to determine solution-based structures of native or mutant SAM domains.

Three dimensional structure may be determined by applying the structural coordinates of a SAM domain structure to other data such as an amino acid sequence, X-ray crystallographic diffraction data, or nuclear magnetic resonance (NMR) data. Homology modeling, molecular replacement, and nuclear magnetic resonance methods using these other data sets are described below.

Homology modeling (also known as comparative modeling or knowledge-based modeling) methods develop a three dimensional model from a polypeptide sequence based on the structures of known proteins. In the present invention the method utilizes a computer representation of the three dimensional structure of a SAM domain, preferably the EphA SAM domain, more preferably the EphA4 SAM domain, or a complex of same, a computer representation of the amino acid sequence of a polypeptide with an unknown structure, and standard computer representations of the structures of amino acids. The method in particular comprises the steps of; (a) identifying structurally conserved and variable regions in the known structure; (b) aligning the amino acid sequences of the known structure and unknown structure (c) generating coordinates of main chain atoms and side chain atoms in structurally conserved and variable regions of the unknown structure based on the coordinates of the known structure thereby obtaining a homology model; and (d) refining the homology model to obtain a three dimensional structure for the unknown structure. This method is well known to those skilled in the art (Greer, 1985, Sceince 228, 1055; Bundell et al 1988, Eur. J. Biochem. 172, 513; Knighton et al., 1992, Science 258:130-135.

http://biochem.vt.edu/courses/modeling/homology.htn). Computer programs that can be used in homology modeling are Quanta and the Homology module in the Insight II modeling package distributed by Molecular Simulations Inc, or MODELLER (Rockefeller University, www.iucr.ac.uk/sinris-top/logical/prg-modeller.html).

In step (a) of the homology modeling method, the known SAM domain structure (e.g. structure of the EphA4 SAM domain) is examined to identify the structurally conserved regions (SCRs) from which an average structure, or framework, can be constructed for these regions of the protein. Variable regions (VRs), in which known structures may differ in conformation, also must be identified. SCRs generally correspond to the elements of secondary structure, such as alpha-helices (the four α-helices in the EphA4 SAM domain) and beta-sheets, and to ligand- and substrate-binding sites. The VRs usually lie on the surface of the proteins and form the loops where the main chain turns.

Many methods are available for sequence alignment of known structures and unknown structure. Sequence alignments generally are based on the dynamic programming algorithm of Needleman and Wunsch [J. Mol. Biol. 48: 442-453, 1970]. Current methods include FASTA, Smith-Waterman, and BLASTP, with the BLASTP method differing from the other two in not allowing gaps. Scoring of alignments typically involves construction of a 20x20 matrix in which identical amino acids and those of similar character (i.e., conservative substitutions) may be scored higher than those of different character. Substitution schemes which may be used to score alignments include the scoring matrices PAM (Dayhoff et al., Meth. Enzymol. 91: 524-545, 1983), and BLOSUM (Henikoff and Henikoff, Proc. Nat. Acad. Sci. USA 89: 10915-10919, 1992), and the matrices based on alignments derived from three-dimensional structures including that of Johnson and Overington (JO matrices) (J. Mol. Biol. 233: 716-738, 1993).

Alignment based solely on sequence may be used, though other structural features also may be taken into account. In Quanta, multiple sequence alignment algorithms are available that may be used when aligning a sequence of the unknown with the known structures. Four scoring systems (i.e. sequence homology, secondary structure homology, residue accessibility homology, CA-CA distance homology) are available, each of which may be evaluated during an alignment so that relative statistical weights may be assigned.

When generating coordinates for the unknown structure, main chain atoms and side chain atoms, both in SCRs and VRs need to be modeled. A variety of approaches known to those skilled in the art may be used to assign coordinates to the unknown. In particular, the coordinates of the main chain atoms of SCRs will be transferred to the unknown structure. VRs correspond most often to the loops on the surface of the polypeptide and if a loop in the known structure is a good model for the unknown, then the main chain coordinates of the known structure may be copied. Side chain coordinates of SCRs and VRs are copied if the residue type in the unknown is identical to or very similar to that in the known structure. For other side chain coordinates, a side chain rotamer library may be used to define the side chain coordinates. When a good model for a loop cannot be found fragment databases may be searched for loops in other proteins that may provide a suitable model for

the unknown. If desired, the loop may then be subjected to conformational searching to identify low energy conformers if desired.

Once a homology model has been generated it should be analyzed to determine its correctness. A computer program available to assist in this analysis is the Protein Health module in Quanta which provides a variety of tests. Other programs that provide structure analysis along with output include PROCHECK and 3D-Profiler [Luthy R. et al, Nature 356: 83-85, 1992; and Bowie, J.U. et al, Science 253: 164-170, 1991]. Once any irregularities have been resolved, the entire structure may be further refined. Refinement may consist of energy minimization with restraints, especially for the SCRs. Restraints may be gradually removed for subsequent minimizations. Molecular dynamics may also be applied in conjunction with energy minimization.

Molecular replacement involves applying X-ray diffraction data of a known structure to the incomplete X-ray crystallographic data set of a polypeptide of unknown structure. The method can be used to define the phases describing the X-ray diffraction data of a polypeptide of unknown structure when only the amplitudes are known. Commonly used computer software packages for molecular replacement are X-PLOR (Brunger 1992, Nature 355: 472-475), AMoRE (Navaza, 1994, Acta Crystallogr. A50:157-163), the CCP4 package (Collaborative Computational Project, Number 4, "The CCP4 Suite: Programs for Protein Crystallography", Acta Cryst., Vol. D50, pp. 760-763, 1994), and the MERLOT package (P.M.D. Fitzgerald, J. Appl. Cryst., Vol. 21, pp. 273-278, 1988). It is preferable that the resulting structure not exhibit a root-mean-square deviation of more than 3 Å.

The objective of molecular replacement is to align positions of atoms in the unit cell by matching electron diffraction data from two crystals. Molecular replacement computer programs generally involve the following steps: (1) determining the number of molecules in the unit cell and defining the angles between them; (2) rotating the diffraction data to define the orientation of the molecules in the unit cell; (3) translating the electron density in three dimensions to correctly position the molecules in the unit cell; (4) determining the amplitudes and phases of the X-ray diffraction data and calculating an R-factor calculated from the reference data set and from the new data wherein an R-factor between 30-50% indicates that the orientations of the atoms in the unit cell have been reasonably determined by the method; and (5) optionally decreasing the R-factor to about 20% by refining the new electron density map using iterative refinement techniques known to those skilled in the art.

In an embodiment of the invention, a method is provided for determining three dimensional structures of polypeptides with unknown structure by applying the structural coordinates of a SAM domain structure to an incomplete X-ray crystallographic data set for a polypeptide of unknown structure, and determining a low energy conformation of the resulting structure.

The structural coordinates of a SAM domain structure may be applied to nuclear magnetic resonance (NMR) data to determine the three dimensional structures of polypeptides. (See for example, Wuthrich, 1986, John Wiley and Sons, New York: 176-199; Pflugrath et al., 1986, J. Molecular Biology 189: 383-386; Kline et al., 1986 J. Molecular Biology 189:377-382). While the secondary structure of a polypeptide may often be determined by NMR data, the spatial connections

between individual pieces of secondary structure are not as readily determined. The structural coordinates of a polypeptide defined by X-ray crystallography can guide the NMR spectroscopist to an understanding of the spatial interactions between secondary structural elements in a polypeptide of related structure. Information on spatial interactions between secondary structural elements can greatly simplify Nuclear Overhauser Effect (NOE) data from two-dimensional NMR experiments. In addition, applying the structural coordinates after the determination of secondary structure by NMR techniques simplifies the assignment of NOE's relating to particular amino acids in the polypeptide sequence and does not greatly bias the NMR analysis of polypeptide structure.

In an embodiment, the invention relates to a method of determining three dimensional structures of polypeptides with unknown structures by applying the structural coordinates of a SAM domain structure to nuclear magnetic resonance (NMR) data of the unknown structure. This method comprises the steps of: (a) determining the secondary structure of an unknown structure using NMR data; and (b) simplifying the assignment of through-space interactions of amino acids. The term " through-space interactions" defines the orientation of the secondary structural elements in the three dimensional structure and the distances between amino acids from different portions of the amino acid sequence. The term "assignment" defines a method of analyzing NMR data and identifying which amino acids give rise to signals in the NMR spectrum.

Identification of Potential Modulators of SAM Domains

Modulators of a SAM domain may be designed and identified that may modify the inappropriate activity of a SAM domain involved in a clinical disorder. The rational design and identification of modulators of SAM domains can be accomplished by utilizing the atomic structural coordinates that define a SAM domain's three dimensional structure.

Modulators may include substances that bind to or mimic the residues of a SAM domain that are required for dimerization of SAM domains. For example, a substance that binds to or mimics the interface residues of an EphA SAM domain (e.g. Val 913, Val 914, Met 972, Met 976, Met 979, Val 944, and Leu 940), or the proximal residues of an EphA SAM domain (e.g. Ile 959 to Lys) may modify inappropriate activity of a SAM domain involved in a clinical disorder.

Structure-based modulator design identification methods are powerful techniques that can involve searches of computer databases containing a variety of potential modulators and chemical functional groups. (See Kuntz et al., 1994, Acc. Chem. Res. 27:117; Guida, 1994, Current Opinion in Struc. Biol. 4: 777; and Colman, 1994, Current Opinion in Struc. Biol. 4: 868, for reviews of structure-based drug design and identification;and Kuntz et al 1982, J. Mol. Biol. 162:269; Kuntz et al., 1994, Acc. Chem. Res. 27: 117; Meng et al., 1992, J. Compt. Chem. 13: 505; Bohm, 1994, J. Comp. Aided Molec. Design 8: 623 for methods of structure-based modulator design).

The SAM domain three dimensional structure described herein, and the three dimensional structures of other polypeptides determined by the homology modeling, molecular replacement, and NMR techniques described herein can also be applied to modulator design and identification methods.

Modulators of SAM domains may be identified by docking the computer representation of compounds from a database of molecules. Databases which may be used include ACD (Molecular Designs Limited), NCI (National Cancer Institute), CCDC (Cambridge Crystallographic Data Center), CAST (Chemical Abstract Service), Derwent (Derwent Information Limited), Maybridge (Maybridge Chemical Company Ltd), Aldrich (Aldrich Chemical Company), DOCK (University of California in San Francisco), and the Directory of Natural Products (Chapman & Hall). Computer programs such as CONCORD (Tripos Associates) or DB-Converter (Molecular Simulations Limited) can be used to convert a data set represented in two dimensions to one represented in three dimensions.

Generally, the computer programs comprise the following steps:

(a) docking the structure of a compound into an active-site of a polypeptide (e.g.. EphA4 SAM domain) using the computer program, or by interactively moving the compound into the active-site;

(b) characterizing the geometry and the complementary interactions formed between the atoms of the active-site and the compound; and optionally

(c) searching libraries for molecular fragments which can fit into the empty space between the compound and active site and can be linked to the compound; and

(d) linking the fragments found in (c) to the compound and evaluating the new modified compound.

"Docking" refers to a process of placing a compound in close proximity with an active site of a polypeptide (e.g.. an Eph SAM domain), or a process of finding low energy conformations of a compound/polypeptide complex (e.g. compound/Eph SAM domain).

Examples of other computer programs that may be used for structure-based modulator design are CAVEAT (Bartlett et al., 1989, in "Chemical and Biological Problems in Molecular Recognition", Roberts, S.M. Ley, S.V.; Campbell, N.M. eds; Royal Society of Chemistry: Cambridge, pp 182-196); FLOG (Miller et al., 1994, J. Comp. Aided Molec. Design 8:153); PRO Modulator (Clark et al., 1995 J. Comp. Aided Molec. Design 9:13); MCSS (Miranker and Karplus, 1991, Proteins: Structure, Fuction, and Genetics 8:195); and, GRID (Goodford, 1985, J. Med. Chem. 28:849).

In an embodiment of the invention, a method is provided for identifying potential modulators of SAM domain function. The method utilizes the structural coordinates of a SAM domain three dimensional structure. The method comprises the steps of (a) removing a computer representation of a SAM domain structure, preferably an Eph SAM domain structure, more preferably an EphA4 SAM domain structure, and docking a computer representation of a compound from a computer data base with a computer representation of the active site of the SAM domain; (b) determining a conformation of the complex with a favourable geometric fit or favorable complementary interactions; and (c) identifying compounds that best fit the SAM domain active-site as potential modulators of SAM domain function. The initial SAM domain structure may or may not have compounds bound to it. A favourable geometric fit occurs when the surface areas of a

compound in a compound-SAM domain complex is in close proximity with the surface area of the active-site of the SAM domain without forming unfavorable interactions. A favourable complementary interaction occurs where a compound in a compound-SAM domain complex interacts by hydrophobic, aromatic, ionic, or hydrogen donating and accepting forces, with the active-site of a SAM domain without forming unfavorable interactions. Unfavourable interactions may be steric hindrance between atoms in the compound and atoms in the SAM active-site.

In another embodiment, potential modulators are identified utilizing a three dimensional structure of a SAM domain with or without compounds bound to it. The method comprises the steps of (a) modifying a computer representation of a SAM domain (e.g. an Eph SAM domain) having one or more compounds bound to it, where the computer representations of the compound or compounds and SAM domain are defined by atomic structural coordinates; (b) determining a conformation of the complex with a favorable geometric fit and favorable complementary interactions; and (c) identifying the compounds that best fit the SAM active site as potential modulators. A computer representation may be modified by deleting or adding a chemical group or groups. Computer representations of the chemical groups can be selected from a computer database.

Another way of identifying potential modulators is to modify an existing modulator in the polypeptide active-site. The computer representation of modulators can be modified within the computer representation of a SAM domain active-site. This technique is described in detail in Molecular Simulations User Manual, 1995 in LUDL The computer representation of a modulator may be modified by deleting a chemical group or groups, or by adding a chemical group or groups. After each modification to a compound, the atoms of the modified compound and active-site can be shifted in conformation and the distance between the modulator and the active site atoms may be scored on the basis of geometric fit and favourable complementary interactions between the molecules. Compounds with favourable scores are potential modulators.

Compounds designed by modulator building or modulator searching computer programs may be screened to identify potential modulators. Examples of such computer programs include programs in the Molecular Simulations Package (Catalyst), ISIS/HOST, ISIS/BASE, and ISIS/DRAW (Molecular Designs Limited), and UNITY (Tripos Associates). A building program may be used to replace computer representations of chemical groups in a compound complexed with a SAM domain with groups from a computer data base. A searching program may be used to search computer representations of compounds from a computer database that have similar three dimensional structures and similar chemical groups as a compound that binds to a SAM domain. The programs may be operated on the structure of the active-site of the three dimensional structure of an Eph SAM domain, preferably an EphA4 SAM domain.

A typical program may comprise the following steps:

(a) mapping chemical features of the compound such as by hydrogen bond donors or acceptors, hydrophobic/lipophilic sites, positively ionizable sites, or negatively ionizable sites;

(b) adding geometric constraints to selected mapped features;

(c) searching data bases with the model generated in (b).

In an embodiment of the invention a method of identifying potential modulators of a SAM domain, preferably an Eph SAM domain, more preferably an EphA SAM domain, is provided using the three dimensional conformation of the SAM domain in various modulator construction or modulator searching computer programs on compounds complexed with the SAM domain. The method comprises the steps of (a) removing a computer representation of one or more compounds complexed with a SAM domain; (b) (i) searching a data base for a compound with a similar geometric structure or similar chemical groups to the removed compounds using a computer program that searches computer representations of compounds from a database that have similar three dimensional structures and similar chemical groups, or (ii) replacing portions of the compounds complexed with the SAM domain with similar chemical structures (i.e. nearly identical shape and volume) from a database using a compound construction computer program that replaces computer representations of chemical groups with groups from a computer database, where the representations of the compounds are defined by structural coordinates.

Potential modulators of SAM domains identified using the above-described methods may be prepared using methods described in standard reference sources utilized by those skilled in the art. For example, organic compounds may be prepared by organic synthetic methods described in references such as March, 1994 Advanced Organic Chemistry: Reactions, Mechanisms, and Structure, New York, McGraw Hill.

Cellular assays, as well as animal model assays *in vivo*, may be used to test the activity of a potential modulator of a SAM domain as well as diagnose a disease associated with inappropriate SAM domain activity. *In vivo* assays are also useful for testing the bioactivity of a potential modulator designed by the methods of the invention.

The invention also relates to a potential modulator identified by the methods of the invention.

Peptides

The invention provides peptide molecules that modulate SAM domain function. The molecules are derived from the interface residues necessary for dimer formation. For example, peptides of the invention include the amino acids Val 913, Val 914, Met 972, Met 976, Met 979, Val 944, and Leu 940 of the EphA4 SAM domain. Other proteins containing sequences corresponding to the sequences necessary for dimer formation of a SAM domain may be identified with a protein homology search, for example by searching available databases such as GenBank or SwissProt and various search algorithms and/or programs may be used including FASTA, BLAST (available as a part of the GCG sequence analysis package, University of Wisconsin, Madison, Wis.), or ENTREZ (National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD).

In accordance with an embodiment of the invention, specific peptides are contemplated that mediate SAM domain function comprising VVSV (SEQ ID. NO. 21), SAVVSV (SEQ ID. NO.22), FSAVV (SEQ ID. NO.23), FSAVVSV (SEQ ID. NO. 24), FSAVVSVGD (SEQ ID. NO. 25),

VVSVGDWL (SEQ ID. NO. 26), FNTV (SEQ ID. NO. 27), FNTVDE (SEQ ID. NO. 28), FNTVDEWL (SEQ ID. NO. 29), TSFNTVDEWL (SEQ ID. NO. 30), TSFNTV (SEQ ID. NO. 31), YTSFNTV (SEQ ID. NO. 32), RSEV (SEQ ID. NO. 33), RSEVLG (SEQ ID. NO. 34), RSEVLGWD (SEQ ID. NO. 35), VPFRSEV (SEQ ID. NO. 36), and VPFRSEVLGW (SEQ ID. NO. 37).

In accordance with another embodiment of the invention, specific peptides are contemplated that mediate SAM domain function. In particular, a peptide of the formula I is provided which mediates SAM domain function:

$$X\text{-}X^1\text{-}X^2\text{-}X^3\text{-}X^4\text{-}X^5\text{-}X^6 \qquad I$$

wherein X and X^6 represent 0 to 70, preferably 0 to 50 amino acids, more preferably 2 to 20 amino acids, and X^1 represents Leu, Phe, Asp, Ala, Glu, or Gly, preferably Leu or Gly, X^2 represents Glu, Asp, Ser, Ile, Ala, Arg, Lys, and Gln, preferably Glu or Asp, X^3 represents Ala, Val, Glu, Phe, Ser, Ile, Met, Leu, His, Gln, Arg, or Asp preferably Ala, Val, or Phe, X^4 is Val, Leu, Met, Phe, and Ile, preferably Val or Leu, or Phe, X^5 is Val, Ser, Leu, Asp, Ala, Pro, Asn, Lys, or Cys, preferably Val or Ser.

In an embodiment of the present invention a peptide of the formula I is provided: wherein X represents TT, ID, TS, DD, GYTT (SEQ ID. NO. 38), AAGYTT (SEQ ID. NO. 39), FTAAGYTT (SEQ ID. NO. 40), DNFTAAGYTT (SEQ ID. NO. 41), or YKDNFTAAGYTT (SEQ ID. NO. 42). In another embodiment X^6 represents HM, HMSQ (SEQ ID. NO. 43), HMSQD (SEQ ID. NO. 44), HMSQDD (SEQ ID. NO. 45), HMSQDDLA (SEQ ID. NO. 46), QMMM (SEQ ID. NO. 47), QMMMED (SEQ ID. NO. 48), QMMMEDLL (SEQ ID. NO. 49), DITE (SEQ ID. NO. 50), DITEED (SEQ ID. NO. 51), DITEEDL (SEQ ID. NO. 52), NLTE (SEQ ID. NO. 53), NLTEND (SEQ ID. NO. 54), NLTENDI (SEQ ID. NO. 55).

Preferred peptides of the formula I include the following: X-LEAVV-X^6, X-FDVVS-X^6, X-LEFLS-X^6, X-GARFL-X^6, LEAVV (SEQ ID. NO. 56), TTLEAVV (SEQ ID. NO. 57), LEAVVHM (SEQ ID. NO. 58), LEAVVHMSQ (SEQ ID. NO. 59), LEAVVHMSQD (SEQ ID. NO. 60), LEAVVHMSQDDL (SEQ ID. NO. 61), LEAVVHMSQDDLAR (SEQ ID. NO. 62), TTLEAVVHMS (SEQ ID. NO. 63), TTLEAVVHMSQD (SEQ ID. NO. 64), TTLEAVVHMSQDDL (SEQ ID. NO. 65), TTLEAVVHMSQDDLAR (SEQ ID. NO. 66), GYTTLEAVV (SEQ ID. NO. 67), GYTTLEAVVHMS (SEQ ID. NO. 68), GYTTLEAVVHMSQD (SEQ ID. NO. 69), GYTTLEAVVHMSQDDL (SEQ ID. NO. 70), GYTTLEAVVHMSQDDLAR (SEQ ID. NO. 71), FDVVS (SEQ ID. NO. 72), FDVVSQ (SEQ ID. NO. 73), FDVVSQMM (SEQ ID. NO. 74), FDVVSQMMME (SEQ ID. NO. 75), FDVVSQMMMEDIL (SEQ ID. NO. 76), TSFDVVS (SEQ ID. NO. 77), TSFDVVSQ (SEQ ID. NO. 78), TSFDVVSQMM (SEQ ID. NO. 79), TSFDVVSQMMME (SEQ ID. NO. 80), TSFDVVSQMMMEDIL (SEQ ID. NO. 81), LEFLS (SEQ ID. NO. 82), LEFLSD (SEQ ID. NO. 83), LEFLSDIT (SEQ ID. NO. 84), LEFLSDITEE (SEQ ID. NO. 85), LEFLSDITEEDL (SEQ ID. NO. 86), DDLEFLS (SEQ ID. NO. 87), GWDDLEFLS (SEQ ID. NO. 88), DDLEFLSD (SEQ ID. NO. 89), DDLEFLSDIT (SEQ ID. NO. 90), DDLEFLSDITEE (SEQ ID.

NO. 91), DDLEFLSDITEEDL (SEQ ID. NO. 92), GARFL (SEQ ID. NO. 93), GARFLN (SEQ ID. NO. 94), GARFLNLT (SEQ ID. NO. 95), GARFLNLTEN (SEQ ID. NO. 96), and IDGARFL (SEQ ID. NO. 97).

In accordance with another embodiment of the invention, specific peptides are contemplated that mediate SAM domain function. In particular, a peptide of the formula II is provided which mediates SAM domain function:

$$X^7\text{-}X^8\text{-}X^9\text{-}X^{10}\text{-}X^{11}\text{-}X^{12}\text{-}X^{13}\text{-}X^{14}\text{-}X^{15}\text{-}X^{16} \qquad \text{II}$$

wherein X^7 and X^{16} represent 0 to 70, preferably 0 to 50 amino acids, more preferably 2 to 20 amino acids, and X^8 represents Met, Ile, Ser, Leu, Asn, Phe, or Val, preferably Met, X^9 represents Arg, Ser, Lys, Met, Leu, Glu, Gln, or Asn, preferably Gln or Arg, X^{10} represents Thr, Ala, Arg, Leu, Ser, Glu, Asp, Met, Lys, Gln, or Gly, preferably Thr, Ala, or Glu, X^{11} represents Gln, Ser, Glu, Leu, Phe, Asp, Thr, Arg, preferably Gln or Arg, X^{12} represents Met, Ala, Ile, Asn, Ser, Arg, Thr, Pro, Leu, Gln, Val, Lys, preferably Met or Arg, X^{13} represents Gln, Asn, Pro, Ser, Tyr, Glu, Leu, Arg, or Lys, preferably Gln, Asn, or Arg, X^{14} represents Gln, Ala, Pro, Asp, Leu, Lys, Ile, Glu, Arg, or Asn, preferably Gln or Ile, and X^{15} represents Met, Ile, Val, His, Ser, Arg, Lys, Phe, Cys, Glu, Tyr, Ala, Ile, Trp, or Leu.

In an embodiment of the present invention a peptide of the formula II is provided: wherein X^7 represents QA, QV, NK, SVQA (SEQ ID. NO. 98), LSSVQA (SEQ ID. NO. 99), ILSSVQA (SEQ ID. NO. 100), NKILSSVQA (SEQ ID. NO. 101), HQNKILSSVQA (SEQ ID. NO. 102), THQNKILSSVQA (SEQ ID. NO. 103), ENIK (SEQ ID. NO. 104), SQEINK (SEQ ID. NO. 105), KLSQEINK (SEQ ID. NO. 106), ILNSIQV (SEQ ID. NO. 107), or NSIQV (SEQ ID. NO. 108). In another embodiment X^7 is HG, QS, HGRM (SEQ ID. NO. 109), HGRMVP (SEQ ID. NO. 110), QSVEV (SEQ ID. NO. 111), or TRKP (SEQ ID. NO. 112).

Preferred peptides of the formula II include the following: X^7-MRTQMQQM-X^{16}, X^7-MRAQMNQI-X^{16}, X^7-NEERRSIF-X^{16}, MRTQMQQM (SEQ ID. NO. 113), QAMRTQMQQM (SEQ ID. NO. 114), SVQAMRTQMQQM (SEQ ID. NO. 115), LSSVQAMRTQMQQM (SEQ ID. NO. 116), ILSSVQAMRTQMQQM (SEQ ID. NO. 117), MRTQMQQMHG (SEQ ID. NO. 118), MRTQMQQMHGRM (SEQ ID. NO. 119), MRTQMQQMHGRMVPV (SEQ ID. NO. 120), NEERRSIF (SEQ ID. NO. 121), INKNEERRSIF (SEQ ID. NO. 122), NEERRSIFTRKP (SEQ ID. NO. 123). MRAQMNQI (SEQ ID. NO. 124), MRAQMNQIQS (SEQ ID. NO. 125), MRAQMNQIQSVEV (SEQ ID. NO. 126).

All of the peptides of the invention, as well as molecules substantially homologous, complementary or otherwise functionally or structurally equivalent to these peptides may be used for purposes of the present invention. In addition to full-length peptides of the invention, truncations of the peptides are contemplated in the present invention. Truncated peptides may comprise peptides of about 7 to 10 amino acid residues

The truncated peptides may have an amino group (-NH2), a hydrophobic group (for example, carbobenzoxyl, dansyl, or T-butyloxycarbonyl), an acetyl group, a 9-fluorenylmethoxy-

carbonyl (PMOC) group, or a macromolecule including but not limited to lipid-fatty acid conjugates, polyethylene glycol, or carbohydrates at the amino terminal end. The truncated peptides may have a carboxyl group, an amido group, a T-butyloxycarbonyl group, or a macromolecule including but not limited to lipid-fatty acid conjugates, polyethylene glycol, or carbohydrates at the carboxy terminal end.

The peptides of the invention may also include analogs of a peptide of the invention and/or truncations of the peptide, which may include, but are not limited to a peptide of the invention containing one or more amino acid insertions, additions, or deletions, or both. Analogs of the peptide of the invention exhibit the activity characteristic of the peptide e.g. interference with SAM domain dimer formation, and may further possess additional advantageous features such as increased bioavailability, stability, or reduced host immune recognition. One or more amino acid insertions may be introduced into a peptide of the invention. Amino acid insertions may consist of a single amino acid residue or sequential amino acids.

One or more amino acids, preferably one to five amino acids, may be added to the right or left termini of a peptide of the invention. Deletions may consist of the removal of one or more amino acids, or discrete portions from the peptide sequence. The deleted amino acids may or may not be contiguous. The lower limit length of the resulting analog with a deletion mutation is about 7 amino acids.

The invention also includes a peptide conjugated with a selected protein, or a selectable marker (see below) to produce fusion proteins.

The peptides of the invention may be prepared using recombinant DNA methods. Accordingly, nucleic acid molecules which encode a peptide of the invention may be incorporated in a known manner into an appropriate expression vector which ensures good expression of the peptide. Possible expression vectors include but are not limited to cosmids, plasmids, or modified viruses so long as the vector is compatible with the host cell used. The expression vectors contain a nucleic acid molecule encoding a peptide of the invention and the necessary regulatory sequences for the transcription and translation of the inserted protein-sequence. Suitable regulatory sequences may be obtained from a variety of sources, including bacterial, fungal, viral, mammalian, or insect genes. (For example, see the regulatory sequences described in Goeddel, Gene Expression Technology: Methods in Enzymology 185, Academic Press, San Diego, CA (1990). Selection of appropriate regulatory sequences is dependent on the host cell chosen, and may be readily accomplished by one of ordinary skill in the art. Other sequences, such as an origin of replication, additional DNA restriction sites, enhancers, and sequences conferring inducibility of transcription may also be incorporated into the expression vector.

The recombinant expression vectors may also contain a selectable marker gene which facilitates the selection of transformed or transfected host cells. Suitable selectable marker genes are genes encoding proteins such as G418 and hygromycin which confer resistance to certain drugs, β-galactosidase, chloramphenicol acetyltransferase, firefly luciferase, or an immunoglobulin or portion

thereof such as the Fc portion of an immunoglobulin preferably IgG. The selectable markers may be introduced on a separate vector from the nucleic acid of interest.

The recombinant expression vectors may also contain genes that encode a fusion portion which provides increased expression of the recombinant peptide; increased solubility of the recombinant peptide; and/or aid in the purification of the recombinant peptide by acting as a ligand in affinity purification. For example, a proteolytic cleavage site may be inserted in the recombinant peptide to allow separation of the recombinant peptide from the fusion portion after purification of the fusion protein. Examples of fusion expression vectors include pGEX (Amrad Corp., Melbourne, Australia), pMAL (New England Biolabs, Beverly, MA) and pRIT5 (Pharmacia, Piscataway, NJ) which fuse glutathione S-transferase (GST), maltose E binding protein, or protein A, respectively, to the recombinant protein.

Recombinant expression vectors may be introduced into host cells to produce a transformant host cell. Transformant host cells include prokaryotic and eukaryotic cells which have been transformed or transfected with a recombinant expression vector of the invention. The terms "transformed with", "transfected with", "transformation" and "transfection" are intended to include the introduction of nucleic acid (e.g. a vector) into a cell by one of many techniques known in the art. For example, prokaryotic cells can be transformed with nucleic acid by electroporation or calcium-chloride mediated transformation. Nucleic acid can be introduced into mammalian cells using conventional techniques such as calcium phosphate or calcium chloride co-precipitation, DEAE-dextran-mediated transfection, lipofectin, electroporation or microinjection. Suitable methods for transforming and transfecting host cells may be found in Sambrook et al. (Molecular Cloning: A Laboratory Manual, 2nd Edition, Cold Spring Harbor Laboratory press (1989)), and other laboratory textbooks.

Suitable host cells include a wide variety of prokaryotic and eukaryotic host cells. For example, the peptides of the invention may be expressed in bacterial cells such as *E. coli*, insect cells (using baculovirus), yeast cells or mammalian cells. Other suitable host cells can be found in Goeddel, Gene Expression Technology: Methods in Enzymology 185, Academic Press, San Diego, CA (1991).

The peptides of the invention may be tyrosine phosphorylated using the method described in Reedijk et al. (The EMBO Journal 11(4):1365, 1992). For example, tyrosine phosphorylation may be induced by infecting bacteria harbouring a plasmid containing a nucleotide sequence encoding a peptide of the invention, with a λgt11 bacteriophage encoding the cytoplasmic domain of the Elk tyrosine kinase as a LacZ-Elk fusion. Bacteria containing the plasmid and bacteriophage as a lysogen are isolated. Following induction of the lysogen, the expressed peptide becomes phosphorylated by the Elk tyrosine kinase.

The peptides of the invention may be synthesized by conventional techniques. For example, the peptides may be synthesized by chemical synthesis using solid phase peptide synthesis. These methods employ either solid or solution phase synthesis methods (see for example, J. M. Stewart, and J.D. Young, Solid Phase Peptide Synthesis, 2nd Ed., Pierce Chemical Co., Rockford Ill. (1984) and G.

Barany and R.B. Merrifield, The Peptides: Analysis Synthesis, Biology editors E. Gross and J. Meienhofer Vol. 2 Academic Press, New York, 1980, pp. 3-254 for solid phase synthesis techniques; and M Bodansky, Principles fo Peptide Synthesis, Springer-Verlag, Berlin 1984, and E. Gross and J. Meienhofer, Eds., The Peptides: Analysis, Synthesis, Biologu, suprs, Vol 1, for classical solution synthesis). By way of example, the peptides may be synthesized using 9-fluorenyl methoxycarbonyl (Fmoc) solid phase chemistry with direct incorporation of phosphotyrosine as the N-fluorenylmethoxy-carbonyl-O-dimethyl phosphono-L-tyrosine derivative.

N-terminal or C-terminal fusion proteins comprising a peptide of the invention conjugated with other molecules may be prepared by fusing, through recombinant techniques, the N-terminal or C-terminal of the peptide, and the sequence of a selected protein or selectable marker with a desired biological function. The resultant fusion proteins contain the peptide fused to the selected protein or marker protein as described herein. Examples of proteins which may be used to prepare fusion proteins include immunoglobulins, glutathione-S-transferase (GST), hemagglutinin (HA), and truncated myc.

Cyclic derivatives of the peptides of the invention are also part of the present invention. Cyclization may allow the peptide to assume a more favorable conformation for association with molecules in complexes of the invention. Cyclization may be achieved using techniques known in the art. For example, disulfide bonds may be formed between two appropriately spaced components having free sulfhydryl groups, or an amide bond may be formed between an amino group of one component and a carboxyl group of another component. Cyclization may also be achieved using an azobenzene-containing amino acid as described by Ulysse, L., et al., J. Am. Chem. Soc. 1995, 117, 8466-8467. The side chains of Tyr and Asn may be linked to form cyclic peptides. The components that form the bonds may be side chains of amino acids, non-amino acid components or a combination of the two. In an embodiment of the invention, cyclic peptides are contemplated that have a beta-turn in the right position. Beta-turns may be introduced into the peptides of the invention by adding the amino acids Pro-Gly at the right position.

It may be desirable to produce a cyclic peptide that is more flexible than the cyclic peptides containing peptide bond linkages as described above. A more flexible peptide may be prepared by introducing cysteines at the right and left position of the peptide and forming a disulphide bridge between the two cysteines. The two cysteines are arranged so as not to deform the beta-sheet and turn. The peptide is more flexible as a result of the length of the disulfide linkage and the smaller number of hydrogen bonds in the beta-sheet portion. The relative flexibility of a cyclic peptide can be determined by molecular dynamics simulations. Peptide mimetics may be designed based on information obtained by systematic replacement of L-amino acids by D-amino acids, replacement of side chains with groups having different electronic properties, and by systematic replacement of peptide bonds with amide bond replacements. Local conformational constraints can also be introduced to determine conformational requirements for activity of a candidate peptide mimetic. The mimetics may include isosteric amide bonds, or D-amino acids to stabilize or promote reverse turn conformations and to help stabilize the molecule. Cyclic amino acid analogues may be used to

constrain amino acid residues to particular conformational states. The mimetics can also include mimics of inhibitor peptide secondary structures. These structures can model the 3-dimensional orientation of amino acid residues into the known secondary conformations of the proteins. Peptoids may also be used which are oligomers of N-substituted amino acids and can be used as motifs for the generation of chemically diverse libraries of novel molecules.

Peptides of the invention may be developed using a biological expression system. The use of these systems allows the production of large libraries of random peptide sequences and the screening of these libraries for peptide sequences that interact with particular amino acid residues. Libraries may be produced by cloning synthetic DNA that encodes random peptide sequences into appropriate expression vectors. (see Christian et al 1992, J. Mol. Biol. 227:711; Devlin et al, 1990 Science 249:404; Cwirla et al 1990, Proc. Natl. Acad, Sci. USA, 87:6378). Libraries may also be constructed by concurrent synthesis of overlapping peptides (see U.S. Pat. No. 4,708,871).

Peptides of the invention may be used to identify lead compounds for drug development. The structure of the peptides described herein can be readily determined by a number of methods such as NMR and X-ray crystallography. A comparison of the structures of peptides similar in sequence, but differing in the biological activities they elicit in target molecules can provide information about the structure-activity relationship of the target. Information obtained from the examination of structure-activity relationships can be used to design either modified peptides, or other small molecules or lead compounds which can be tested for predicted properties as related to the target molecule. The activity of the lead compounds can be evaluated using assays similar to those described herein.

Information about structure-activity relationships may also be obtained from co-crystallization studies. In these studies, a peptide with a desired activity is crystallized in association with a target molecule i.e. SAM domain, and the X-ray structure of the complex is determined. The structure can then be compared to the structure of the target molecule in its native state, and information from such a comparison may be used to design compounds expected to possess desired activities.

The peptides of the invention may be converted into pharmaceutical salts by reacting with inorganic acids such as hydrochloric acid, sulfuric acid, hydrobromic acid, phosphoric acid, etc., or organic acids such as formic acid, acetic acid, propionic acid, glycolic acid, lactic acid, pyruvic acid, oxalic acid, succinic acid, malic acid, tartaric acid, citric acid, benzoic acid, salicylic acid, benezenesulfonic acid, and toluenesulfonic acids. The peptides of the invention may be used to prepare antibodies. Conventional methods can be used to prepare the antibodies.

The peptides and antibodies specific for the peptides of the invention may be labelled using conventional methods with various enzymes, fluorescent materials, luminescent materials and radioactive materials. Suitable enzymes, fluorescent materials, luminescent materials, and radioactive material are well known to the skilled artisan. Antibodies and labeled antibodies specific for the peptides of the invention may be used to screen for proteins containing SAM domains.

Computer modelling techniques known in the art may also be used to observe the interaction of a peptide of the invention, and truncations and analogs thereof with a SAM domain (for example, Homology Insight II and Discovery available from BioSym/Molecular Simulations, San Diego, California, U.S.A.). If computer modelling indicates a strong interaction, the peptide can be synthesized and tested for its ability to interfere with SAM domain dimer formation.

Compositions and Methods of Treatment

A purified three dimensional SAM domain structure of the invention, the peptides of the invention, and the modulators identified using the methods of the invention may be used to modify the inappropriate activity of a SAM domain involved in a clinical disorder. They may be used in the treatment and diagnosis of disorders associated with aberrant T cell signaling and to modulate telomere function. In particular, they may be useful in methods for therapy of cellular senescence and immortalization controlled by telomere length and telomerase activity, and as selective immunosuppressants (e.g. in organ transplantation). They may also be useful in the treatment of cancers, such as melanoma, ocular melanoma, leukemia, astrocytoma, glioblastoma, lymphoma, glioma, Hodgkin's lymphoma, multiple myeloma, sarcoma, myosarcoma, cholangiocarcinoma, squamous cell carcinoma, CLL, and cancers of the pancreas, breast, brain, prostate, bladder, thyroid, ovary, uterus, testis, kidney, stomach, colon and rectum, particularly leukemia including B-cell leukemia, T-cell leukemia, null-cell leukemia, myelogenous leukemia, and lymphocytic leukemia,

Further, the three dimensional SAM domain structure of the invention, the peptides of the invention, and the modulators identified using the methods of the invention may be used to modulate the biological activity of an Eph receptor or Eph ligand in a cell, including inhibiting or enhancing signal transduction activities of the receptor or ligand, and in particular modulating a pathway in a cell regulated by the ligand or receptor, particularly those pathways involved in neuronal development, axonal migration, pathfinding and regeneration. The three dimensional SAM domain structure of the invention, the peptides of the invention, and modulators identified using the methods of the invention will be useful as pharmaceuticals to modulate axonogenesis, nerve cell interactions and regeneration, to treat conditions such as neurodegenerative diseases and conditions involving trauma and injury to the nervous system, for example Alzheimer's disease, Parkinson's disease, Huntington's disease, demyelinating diseases, such as multiple sclerosis, amyotrophic lateral sclerosis, bacterial and viral infections of the nervous system, deficiency diseases, such as Wernicke's disease and nutritional polyneuropathy, progressive supranuclear palsy, Shy Drager's syndrome, multistem degeneration and olivo ponto cerebellar atrophy, peripheral nerve damage, and trauma and ischemia resulting from stroke.

The present invention thus provides a method for treating cancer (e.g. leukemia), and disorders associated with T cell signaling, modulating telomere function, or affecting neuronal development or regeneration, in a subject comprising administering to a subject an effective amount of a three dimensional SAM domain structure of the invention, a peptide of the invention, or a modulator identified using the methods of the invention. The invention also contemplates a method for stimulating or inhibiting axonogenesis in a subject comprising administering to a subject an

effective amount of a three dimensional SAM domain structure of the invention, a peptide of the invention, or a modulator identified using the methods of the invention.

The invention still further relates to a pharmaceutical composition which comprises a purified three dimensional SAM domain structure of the invention, a peptide of the invention, or a modulator identified using the methods of the invention, and a pharmaceutically acceptable carrier, diluent or excipient. The pharmaceutical compositions may be used to stimulate or inhibit neuronal development, regeneration and axonal migration associated with neurodegenerative conditions, and conditions involving trauma and injury to the nervous system. They may also be used to treat cancer and disorders associated with T cell signaling, and modulate telomere function.

The compositions of the invention are administered to subjects in a biologically compatible form suitable for pharmaceutical administration *in vivo*. By "biologically compatible form suitable for administration *in vivo*" is meant a form of the protein to be administered in which any toxic effects are outweighed by the therapeutic effects of the protein. The term subject is intended to include mammals and includes humans, dogs, cats, mice, rats, and transgenic species thereof. Administration of a therapeutically active amount of the pharmaceutical compositions of the present invention is defined as an amount effective, at dosages and for periods of time necessary to achieve the desired result. For example, a therapeutically active amount of a three dimensional SAM domain structure of the invention, peptides of the invention, or modulators of the invention may vary according to factors such as the condition, age, sex, and weight of the individual. Dosage regimes may be adjusted to provide the optimum therapeutic response. For example, several divided doses may be administered daily or the dose may be proportionally reduced as indicated by the exigencies of the therapeutic situation.

The active compound may be administered in a convenient manner such as by injection (subcutaneous, intravenous, etc.), oral administration, inhalation, transdermal application, or intracerebral administration. In particular embodiments, pharmaceutical compositions of the invention are administered directly to the peripheral or central nervous system, for example by administration intracerebrally.

A pharmaceutical composition of the invention can be administered to a subject in an appropriate carrier or diluent, co-administered with enzyme inhibitors or in an appropriate carrier such as microporous or solid beads or liposomes. The term "pharmaceutically acceptable carrier" as used herein is intended to include diluents such as saline and aqueous buffer solutions. Liposomes include water-in-oil-in-water emulsions as well as conventional liposomes (Strejan et al., (1984) J. Neuroimmunol 7:27). The active compound may also be administered parenterally or intraperitoneally. Dispersions can also be prepared in glycerol, liquid polyethylene glycols, and mixtures thereof and in oils. Under ordinary conditions of storage and use, these preparations may contain a preservative to prevent the growth of microorganisms. Depending on the route of administration, the active compound may be coated to protect the compound from the action of enzymes, acids and other natural conditions which may inactivate the compound.

The pharmaceutical compositions may be administered locally to stimulate axonogenesis and pathfinding, for example the compositions may be administered in areas of local nerve injury or in areas where normal nerve pathway development has not occurred. The pharmaceutical compositions may also be placed in a specific orientation or alignment along a presumptive pathway to stimulate axon pathfinding along that line, for example the pharmaceutical compositions may be incorporated on microcarriers laid down along the pathway. In particular, the pharmaceutical compositions of the invention may be used to stimulate formation of connections between areas of the brain, such as between the two hemispheres or between the thalamus and ventral midbrain. The pharmaceutical compositions may be used to stimulate formation of the medial tract of the anterior commissure or the habenular interpeduncle.

Therapeutic administration of polypeptides may also be accomplished using gene therapy, A nucleic acid including a promoter operatively linked to a heterologous polypeptide may be used to produce high-level expression of the polypeptide in cells transfected with the nucleic acid. DNA or isolated nucleic acids may be introduced into cells of a subject by conventional nucleic acid delivery systems. Suitable delivery systems include liposomes, naked DNA, and receptor-mediated delivery systems, and viral vectors such as retroviruses, herpes viruses, and adenoviruses.

The following non-limiting example is illustrative of the presént invention:

EXAMPLE

The following methods were used to determine the crystal structure of the SAM domain of the Eph receptor isoform A4.

Protein expression, mutagenesis and purification: The SAM domain of the Eph receptor isoform A4 (residues 890 to 981) was expressed in *E. coli* as a GST fusion protein using the pGEX-2T vector (Pharmacia). The Quickchange kit (Stratagene) was used to generate site directed mutants for dimerization analysis and for heavy atom phasing. Protein was purified by affinity chromatography using glutathione Sepharose beads (Pharmacia). Bound protein was eluted by cleavage with thrombin. After concentrating to 10 mM, protein was applied to a Superdex 75 gel filtration column (Pharmacia) for final purification and characterization.

Crystallization and data collection: Hanging drops containing 1µl of 100 mg/ml native or mutant (Glu 941 Cs) protein in 7mM Hepes pH 7.5 were mixed with equal volumes of reservoir buffer containing 100 mM cacodylate pH 6.5, 7% (w/v) PEG 8000, and 20% (v/v) ethylene glycol. Rod like crystals of approximate dimensions 0.05 x 0.05 x 0.2 mm were obtained overnight. The crystals contain one molecule of the EphA4 SAM domain per asymmetric unit, and belong to the space group $P6_4$, (a = b = 77.14 Å, c = 24.37 Å). The solution dimer corresponds to a crystallographic dimer generated from the asymetric unit by a two fold rotation parallel to the unique crystal axis. Crystals were cryo-protected in reservoir buffer enriched to 20% (w/v) PEG 8000 and 20% (v/v) ethylene glycol prior to stream freezing. Heavy atom derivatives were prepared by soaking crystals overnight in 1-10 mM heavy atom solution prepared in cryo-protection buffer.

Native and derivative diffraction data were collected on frozen crystals (108°K) using a Marr 345 imaging plate detector system with an RU200 rotating anode generator (Table 1). Data processing and reduction was carried out with the HKL, DENZO, and SCALEPACK programs.

Single isomorphous replacement (SIR) protein phases were calculated using lead derivative data collected on two separate protein crystals. The heavy atom site was identified by the Patterson search program HASSP [Terwilliger, 1987]. A Glu 941 to Cys site directed mutant of the EphA4 SAM domain construct was employed for mercury derivatization. The heavy atom position of the mercury derivative data, which was collected on three separate crystals, was identified by difference Fourier synthesis. Multiple isomorphous replacement and anomalous scattering (MIRAS) phases, using only the lead derivative anomalous signal, were calculated and iterative rounds of automatic solvent boundary determination/density modification were performed using the PHASES package [Furey, 1990]. The resultant experimental electron density map allowed for the complete tracing of the SAM domain backbone structure.

Model building and Refinement: Model building was performed using O [Jones, 1991]. A starting model comprising approximately 65% of the total structure was refined using XPLOR [Brünger, 1992]. Bulk solvent correction was applied during refinement and simulated annealing protocols were employed. The remaining structure was built into $2F_o$-F_c electron density maps generated with XPLOR. The final refinement statistics are shown in Table 1. The first 20 residues of the SAM domain construct are disordered (residues 890 to 909) and have not been modeled. No amino acid residues occupy disallowed regions of the Ramachandran plot and 94 % occupy the most favored regions.

Results:

The X-ray crystal structure of the SAM domain from the EphA4 receptor tyrosine kinase (Table 1 and 2) was determined. The boundaries of the structure were defined by limited proteolysis and mass-spectrometry. Overall, the structure of the homodimer is oblong and arises from the association of two 'lobster claw' shaped subunits. Each subunit possesses a globular fold consisting of an N-terminal extended strand segment, followed by four short α helices (α1 to α4) and one long C-terminal helix α5 (Figure 2A, 2B, and 2C). The N- and the C-termini are located on one side of the subunit fold, similar to other protein interaction modules with signaling function (SH3, SH2, PH domains etc.) [Kuriyan, 1997]. However, in contrast to these other domains, the termini compose the functional end of the molecule rather than lying opposite to the ligand-binding surface. As shown in Figure 3A and 3B, the N-terminal strand region and the C-terminal helix α5 extend from the subunit core and interdigitate in a pincer like manner with the termini of a second subunit, to form an elaborate dimer interface. In addition to the N- and C-terminal regions, α-helices α1 and α3 contribute side chains to the dimer interface.

The N-terminal strands cross in an anti-parallel manner and project the side chains of Ala 912, Val 913, Val 914 and Phe 910 downward to form one mandible of the 'lobster claw' shaped subunit. The C-terminal helices α5 also cross in an anti-parallel manner with each α-helix projecting the side chains of Met 972, Met 976, and Met 979, upwards to form the second mandible. Together

these side chains compose a hydrophobic core that is fully continuous with those of the individual subunits. Residues bridging the subunit and interface cores include Trp 919, Ala 922 and Ile 923 from helix α1 and Leu 940 and Val 944 from helix α3. Complementing these hydrophobic interactions, the conserved side chain of arginine 973 forms intermolecular electrostatic interactions with the free carboxylate of glycine 981 and a stabilizing charge/helix dipole interaction with the C-terminus of helix α5 (Figure 2C). Additional polar residues located at or in close proximity to the dimer interface include His 980, Gln 975, His 945, Gln 977, Glu 941 and Ser 911.

In order to identify determinants of dimerization and to test that the crystallographic dimer model reflects the solution structure of the EphA4 SAM domain, SAM domain residues, either singly or in combination, were substituted and the behaviour of these mutants was tested using size exclusion chromatography (Figure 4). In agreement with predictions from the crystal structure, mutations involving the interface residues Val 913, Val 914, Met 972, Met 976, Met 979, Val 944, and Leu 940 abolished dimer formation. In contrast, mutation of Val 969 to Ala, which comprises part of the second hydrophobic surface region (Figure 3A and 3B), did not affect dimerization while mutation of the proximal residue Ile 959 to Lys, appeared to disrupt the integrity of the subunit fold. Additionally, mutation of the surface exposed residues Glu 941, Asp 949, and Ser 968 to cysteine, did not disrupt SAM domain dimerization. In summary, the mutagenesis results are consistent with and support the notion that the SAM domain dimer observed in the crystal structure represents a mechanism through which the SAM domain associates in solution.

To investigate whether the dimer model for the Eph receptor SAM domain has more general relevance for SAM domain containing proteins, the predicted locations of residues that are required for the dimerization of SAM domains on other polypeptides were examined. When mutations that map to conserved features of the subunit core and therefore are likely to disrupt the subunit fold are eliminated, a number of informative mutations stand out. For example, the homo- and hetero-typic dimerization of the Polycomb family of transcriptional repressors ph, RAE28 and Scm, is abolished by mutation of two residues predicted to map to the dimer interface [Kyba, 1998]. These residues, Ile 62 and Trp 1 of the ph SAM domain, correspond to the N-terminal strand residue Phe 910 and the α5 helix residue Met 972, respectively, of the EphA4 SAM domain. Both residues are highly conserved amongst the SAM domains and yet are unlikely to affect the individual subunit fold. The mutation of the latter residue (Met 972 to Lys) in the EphA4 SAM domain yields a compact monomer structure (Figure 4). In addition, the hetero-dimerization of the SAM domain containing proteins Byr2p and Ste4p is disrupted by the substitution of Arg 69 with cysteine[Tu, 1997 #25]. This mutation maps to the interface residue Gln 977 of the EphA4 SAM dimer, and is located at the crossing site of the two α5 helices. Taken together, these observations indicate that the dimer structure of the EphA4 SAM domain may reflect a more general mode of SAM domain dimerization.

The crystallographic model for SAM domain dimerization is attractive for a number of reasons. Firstly, in the case of the Eph receptors, the linkers between the SAM and the catalytic domains is short (5 residues of poorly conserved sequence) so that the N-termini of the dimer would have to be oriented in the same direction and in close proximity if the kinase domains of clustered

receptors were to be juxtaposed. The structure shows this to be the case. Secondly, the mechanism of dimerization revealed by the structure could account for the observation that the SAM domain is found at either terminus of signaling proteins. Because the N- and C-terminal ends of the SAM domain compose the dimer interface, the insertion of a SAM domain at an internal site in a polypeptide chain would sterically restrict access to a second SAM domain, especially if the host sequence was itself structured. The solutions to this dilemma would be to place a SAM domain at the end of a protein (as is usually observed), or to surround it with long linker sequences. In this regard the SAM domain differs from modules such as SH2 and SH3 domains, which can readily be located at internal positions in a polypeptide chain since the ligand-binding site is located opposite to the location of the N- and C-termini [Kuriyan, 1997]. Thirdly, in the case of the liprins we have noted three adjacent SAM domains in a region previously shown to mediate liprin hetero-dimerization [Serra-Pages, 1998]. Because the C-termini of the dimerized SAM domain are in close proximity, on the opposite side from the N-termini, a configuration of stacked SAM domains can be readily envisioned.

SAM dimerization may contribute to receptor oligomerization and activation by bringing catalytic elements into proximity for autophosphorylation. The SAM domain may have a direct inhibitory interaction with the kinase domain that can be competed away by dimerization. Alternatively SAM domain mediated dimerization might maintain opposing catalytic domains in a mutually inaccessible, and thus repressed state. The Eph SAM domains might also recruit signaling partners through heteromeric SAM-SAM interactions, or through specific recognition of cytoplasmic proteins by the Eph SAM dimer.

SAM dimerization might be constitutive, but controlled through co-operative or antagonistic interactions with other clustering forces. Dimerization could potentially be controlled by modifications such as tyrosine phosphorylation, and indeed a residue within the SAM domain of the EphB1 receptor can become tyrosine phosphorylated in vivo [Stein, 1996]. Finally, the five residues that lie C-terminal to the Eph SAM domain represent a potential binding site for PDZ domain proteins[Hock, 1998],which might influence the organization of the SAM domain.

The structure of the EphA4 domain reveals a novel mechanism through which modular domains control protein-protein interactions. Since SAM domains are found in cell surface receptors, cytoplasmic signaling proteins, and transcriptional activators and repressors, as well as chimeric human oncoproteins, these results have general implications for understanding the formation of complexes involved in normal and oncogenic signal transduction.

Having illustrated and described the principles of the invention in a preferred embodiment, it should be appreciated to those skilled in the art that the invention can be modified in arrangement and detail without departure from such principles. All modifications coming within the scope of the following claims are claimed.

All publications, patents and patent applications referred to herein are incorporated by reference in their entirety to the same extent as if each individual publication, patent or patent application was specifically and individually indicated to be incorporated by reference in its entirety.

Detailed Description of the Drawings

Figure 1A shows a sequence alignment of SAM domains from selected proteins. Secondary structure is indicated for the SAM domain from the EphA4 receptor tyrosine kinase. Residue numbers for the start of each SAM domain are shown on the left and Genebank accession numbers on the right. Conserved hydrophobic residues are colored green, acidic residues red, basic residues blue, polar residues orange and glycines are colored pink. Residues at the dimer interface shown in Figure 2C are indicated (•). Liprin α1 contains 3 SAM domains designated S1, S2 and S3.

Figure 1B shows a selection of multi-domain proteins containing SAM domain (S) is shown. Domains listed include, tyrosine or serine/threonine kinase catalytic domains, myosin-like domain, F-actin binding domain (F-actin BD), PDZ domain, SH2 domain, inositol phosphatase catalytic domain (inositol p' tase), GTPase activating domain (GAP), DNA-binding domain (DNA-BD) and a transmembrane region (TM).

Figure 2A, 2B, and 2C. Ribbons depiction of the SAM homo-dimer viewed (Figure 2A) down the twofold symmetry axis and (Figure 2B) perpendicular to the symmetry axis. The dimer subunits are coloured red and blue and α-helices are labeled. (Figure 2C) Ribbons stereo view highlighting the dimer interface region. Aromatic, aliphatic, methionine, histidine and arginine interacting side chains are coloured light blue, green, yellow, orange, and blue (see Figure 1A for residue identification). All ribbon diagrams were generated using RIBBONS [Carson, 1991].

Figure 3A, B. Molecular surface and worm representations of the SAM homodimer. The molecular surface of one subunit is shown with hydrophobic (Met, Val, Leu, Ile, Phe,), basic (Arg, Lys) and acidic (Glu, Asp) side chains coloured green, blue and red, respectively. The two perspectives differ by a 90° rotation about the vertical axis. In Figure 3B the twofold rotation axis relating the two subunits of the dimer is shown. The buried surface area of the dimer interface is 1923 Å. All molecular surfaces were generated using GRASP [Nicholls, 1991].

Figure 4. Gel filtration elution profile of wild type and single or double site mutants of the EphA4 receptor SAM domain. Chromatograms correspond to the loading of equivalent concentrations (10 mM) and total volumes (100 µl) of protein on a Superdex-75 gel filtration column (24 ml bed volume). The column was calibrated using Pharmacia low molecular weight standards.

Table 1. Data Collection, Structure Determination and Refinement Statistics

	Resolution (Å)	Reflections total / unique	Completeness (%)	Rsym* (%)	Rderiv+ (%)	Sites (#)	Phasing Power@ (2.9Å)	FOM& (2.9Å)
Native Data (I/sig >-3)	2.0	17787 / 5663	97.5(91.7)	4(30.1)				
MIR analysis								
Pb Acetate 1	2.3	15572 / 3761	98.0 (96.0)	5.9 (23.0)	17.60	1	2.1	0.41
Pb Acetate 2	2.8	20533 / 2113	98.0 (99.0)	7.1 (19.5)	19.00	1	1.8	0.37
Hg Acetate 1	2.6	9368 / 2569	95.6 (85.5)	6.8 (25.9)	11.60	1	1.2	0.31
Hg Acetate 2	2.9	8683 / 1857	95.9 (90.9)	9.4 (25.9)	16.20	1	1.8	0.36
Hg Acetate 3	2.9	5667 / 1928	98.2 (97.0)	10.0 (28.7)	13.60	1	1.7	0.37
MIRAS combined	(Pb anomalous only)							0.74

Refinement		Reflections	Completeness	Rfactor	Rfree#
30-2.0Å	I/sig > 0	5498	94.6 (86.1)	22.9	27.8
30-2.0Å	I/sig > 2	5194	89.4 (72.5)	21.7	27.3

RMSD from ideal	
Bond Length (Å)	0.01
Bond Angle (°)	1.38
Ave. B factor (Å)	22.3 (33.2 for solvent)
RMSD in B factor (Å)	2.01

total nonhydrogen atoms	561
water molecules	53

*$R_{sym} = 100 \times \Sigma_h \Sigma_i |I_{h,i} - \langle I_h \rangle| / \Sigma_h \Sigma_i I_{h,i}$

+$R_{deriv} = \Sigma_h |F_{nat,h} - F_{deriv,h}| / \Sigma_h F_{nat,h}$

#Free R_{factor} was calculated with 5% of the data.

@Phasing Power is RMS ($|F_h|/E$) where the subscript 'h' represents 'heavy-atom' and E is the residual lack of closure.

&FOM (mean figure of merit) = $\langle | \Sigma P(\alpha) e^{i\alpha} / \Sigma P(\alpha) | \rangle$, where α is the phase and $P(\alpha)$ is the phase probability distribution.

Table 2

REMARK EphA4 SAM domain

5	ATOM	1	CB	PHE	910	29.722	33.942	12.780	1.00	34.97
	ATOM	2	CG	PHE	910	28.445	34.752	12.839	1.00	31.67
	ATOM	3	CD1	PHE	910	27.612	34.694	13.956	1.00	30.66
	ATOM	4	CD2	PHE	910	28.067	35.563	11.771	1.00	31.49
	ATOM	5	CE1	PHE	910	26.424	35.428	14.007	1.00	28.32
10	ATOM	6	CE2	PHE	910	26.881	36.301	11.812	1.00	28.34
	ATOM	7	CZ	PHE	910	26.059	36.230	12.931	1.00	28.11
	ATOM	8	C	PHE	910	28.471	32.190	11.520	1.00	38.36
	ATOM	9	O	PHE	910	28.797	32.277	10.336	1.00	39.23
	ATOM	10	HT1	PHE	910	31.128	32.065	11.446	1.00	10.00
15	ATOM	11	HT2	PHE	910	30.595	30.703	12.275	1.00	10.00
	ATOM	12	N	PHE	910	30.779	31.717	12.369	1.00	37.87
	ATOM	13	HT3	PHE	910	31.459	31.917	13.127	1.00	10.00
	ATOM	14	CA	PHE	910	29.489	32.432	12.628	1.00	37.16
	ATOM	15	N	SER	911	27.236	31.915	11.923	1.00	37.36
20	ATOM	16	H	SER	911	27.001	31.871	12.873	1.00	0.00
	ATOM	17	CA	SER	911	26.138	31.684	10.994	1.00	36.43
	ATOM	18	CB	SER	911	25.532	30.288	11.233	1.00	40.15
	ATOM	19	OG	SER	911	24.503	30.015	10.300	1.00	45.27
	ATOM	20	HG	SER	911	24.111	29.153	10.464	1.00	0.00
25	ATOM	21	C	SER	911	25.075	32.757	11.224	1.00	32.88
	ATOM	22	O	SER	911	24.780	33.107	12.365	1.00	30.65
	ATOM	23	N	ALA	912	24.515	33.274	10.130	1.00	29.12
	ATOM	24	H	ALA	912	24.829	32.997	9.247	1.00	10.00
	ATOM	25	CA	ALA	912	23.491	34.300	10.214	1.00	27.10
30	ATOM	26	CB	ALA	912	23.525	35.190	8.991	1.00	28.25
	ATOM	27	C	ALA	912	22.091	33.725	10.416	1.00	25.69
	ATOM	28	O	ALA	912	21.124	34.480	10.538	1.00	25.68
	ATOM	29	N	VAL	913	21.968	32.395	10.409	1.00	25.64
	ATOM	30	H	VAL	913	22.774	31.852	10.334	1.00	10.00
35	ATOM	31	CA	VAL	913	20.667	31.749	10.602	1.00	25.94
	ATOM	32	CB	VAL	913	20.410	30.608	9.572	1.00	27.20
	ATOM	33	CG1	VAL	913	19.853	31.180	8.278	1.00	28.69
	ATOM	34	CG2	VAL	913	21.683	29.825	9.304	1.00	28.91
	ATOM	35	C	VAL	913	20.519	31.194	12.013	1.00	24.45
40	ATOM	36	O	VAL	913	21.515	30.851	12.659	1.00	24.44
	ATOM	37	N	VAL	914	19.277	31.120	12.486	1.00	22.36
	ATOM	38	H	VAL	914	18.546	31.390	11.893	1.00	10.00
	ATOM	39	CA	VAL	914	18.990	30.618	13.826	1.00	23.06
	ATOM	40	CB	VAL	914	17.612	31.113	14.354	1.00	24.08
45	ATOM	41	CG1	VAL	914	17.600	32.633	14.442	1.00	24.61
	ATOM	42	CG2	VAL	914	16.479	30.620	13.456	1.00	24.73
	ATOM	43	C	VAL	914	19.050	29.096	13.891	1.00	23.11
	ATOM	44	O	VAL	914	19.208	28.427	12.867	1.00	22.16
	ATOM	45	N	SER	915	18.928	28.561	15.104	1.00	24.03
50	ATOM	46	H	SER	915	18.786	29.134	15.882	1.00	10.00
	ATOM	47	CA	SER	915	18.986	27.119	15.340	1.00	24.05
	ATOM	48	CB	SER	915	19.224	26.857	16.827	1.00	23.17
	ATOM	49	OG	SER	915	18.136	27.348	17.596	1.00	24.34
	ATOM	50	HG	SER	915	17.398	26.778	17.408	1.00	10.00
55	ATOM	51	C	SER	915	17.722	26.379	14.905	1.00	23.08
	ATOM	52	O	SER	915	16.632	26.961	14.867	1.00	21.72
	ATOM	53	N	VAL	916	17.879	25.083	14.636	1.00	22.36
	ATOM	54	H	VAL	916	18.769	24.688	14.711	1.00	10.00

ATOM	55	CA	VAL	916	16.771	24.222	14.240	1.00	21.66
ATOM	56	CB	VAL	916	17.243	22.776	13.978	1.00	20.59
ATOM	57	CG1	VAL	916	16.069	21.892	13.601	1.00	20.62
ATOM	58	CG2	VAL	916	18.279	22.760	12.880	1.00	18.09
ATOM	59	C	VAL	916	15.758	24.206	15.380	1.00	21.55
ATOM	60	O	VAL	916	14.552	24.235	15.145	1.00	22.25
ATOM	61	N	GLY	917	16.265	24.201	16.613	1.00	21.67
ATOM	62	H	GLY	917	17.236	24.168	16.755	1.00	10.00
ATOM	63	CA	GLY	917	15.408	24.200	17.788	1.00	21.14
ATOM	64	C	GLY	917	14.484	25.406	17.849	1.00	21.35
ATOM	65	O	GLY	917	13.297	25.271	18.159	1.00	22.38
ATOM	66	N	ASP	918	15.025	26.582	17.535	1.00	21.54
ATOM	67	H	ASP	918	15.955	26.594	17.261	1.00	10.00
ATOM	68	CA	ASP	918	14.254	27.823	17.544	1.00	20.67
ATOM	69	CB	ASP	918	15.165	29.027	17.298	1.00	24.65
ATOM	70	CG	ASP	918	16.073	29.322	18.471	1.00	28.02
ATOM	71	OD1	ASP	918	17.203	29.794	18.235	1.00	30.66
ATOM	72	OD2	ASP	918	15.656	29.089	19.626	1.00	30.44
ATOM	73	C	ASP	918	13.170	27.788	16.482	1.00	17.85
ATOM	74	O	ASP	918	12.042	28.225	16.716	1.00	16.82
ATOM	75	N	TRP	919	13.536	27.296	15.305	1.00	15.94
ATOM	76	H	TRP	919	14.465	26.996	15.186	1.00	10.00
ATOM	77	CA	TRP	919	12.608	27.180	14.191	1.00	14.95
ATOM	78	CB	TRP	919	13.360	26.724	12.931	1.00	15.34
ATOM	79	CG	TRP	919	12.486	26.125	11.871	1.00	13.90
ATOM	80	CD2	TRP	919	12.470	24.754	11.453	1.00	13.38
ATOM	81	CE2	TRP	919	11.460	24.627	10.478	1.00	11.80
ATOM	82	CE3	TRP	919	13.212	23.622	11.813	1.00	13.78
ATOM	83	CD1	TRP	919	11.519	26.757	11.149	1.00	13.86
ATOM	84	NE1	TRP	919	10.895	25.862	10.312	1.00	13.06
ATOM	85	HE1	TRP	919	10.171	26.089	9.693	1.00	10.00
ATOM	86	CZ2	TRP	919	11.167	23.410	9.856	1.00	12.61
ATOM	87	CZ3	TRP	919	12.920	22.408	11.193	1.00	14.43
ATOM	88	CH2	TRP	919	11.905	22.315	10.226	1.00	11.85
ATOM	89	C	TRP	919	11.473	26.213	14.540	1.00	13.78
ATOM	90	O	TRP	919	10.303	26.526	14.332	1.00	14.59
ATOM	91	N	LEU	920	11.819	25.053	15.093	1.00	13.35
ATOM	92	H	LEU	920	12.765	24.865	15.260	1.00	10.00
ATOM	93	CA	LEU	920	10.820	24.058	15.468	1.00	12.53
ATOM	94	CB	LEU	920	11.493	22.771	15.954	1.00	10.48
ATOM	95	CG	LEU	920	12.110	21.896	14.853	1.00	11.20
ATOM	96	CD1	LEU	920	12.821	20.705	15.470	1.00	12.10
ATOM	97	CD2	LEU	920	11.036	21.425	13.879	1.00	10.00
ATOM	98	C	LEU	920	9.836	24.581	16.511	1.00	12.61
ATOM	99	O	LEU	920	8.635	24.334	16.407	1.00	13.77
ATOM	100	N	GLN	921	10.331	25.303	17.512	1.00	13.95
ATOM	101	H	GLN	921	11.298	25.464	17.569	1.00	10.00
ATOM	102	CA	GLN	921	9.447	25.848	18.538	1.00	16.28
ATOM	103	CB	GLN	921	10.247	26.450	19.694	1.00	16.54
ATOM	104	CG	GLN	921	9.368	26.933	20.840	1.00	16.62
ATOM	105	CD	GLN	921	10.165	27.457	22.006	1.00	16.33
ATOM	106	OE1	GLN	921	11.072	26.794	22.500	1.00	16.95
ATOM	107	NE2	GLN	921	9.820	28.650	22.467	1.00	18.37
ATOM	108	HE21	GLN	921	9.081	29.120	22.029	1.00	10.00
ATOM	109	HE22	GLN	921	10.307	29.002	23.245	1.00	10.00
ATOM	110	C	GLN	921	8.504	26.899	17.948	1.00	15.94
ATOM	111	O	GLN	921	7.327	26.953	18.294	1.00	17.35
ATOM	112	N	ALA	922	9.024	27.698	17.021	1.00	17.97
ATOM	113	H	ALA	922	9.965	27.577	16.767	1.00	10.00

ATOM	114	CA	ALA	922	8.260	28.750	16.363	1.00	16.54
ATOM	115	CB	ALA	922	9.178	29.591	15.504	1.00	16.52
ATOM	116	C	ALA	922	7.107	28.207	15.525	1.00	17.54
ATOM	117	O	ALA	922	6.082	28.875	15.364	1.00	16.79
ATOM	118	N	ILE	923	7.287	27.015	14.962	1.00	16.97
ATOM	119	H	ILE	923	8.139	26.545	15.085	1.00	10.00
ATOM	120	CA	ILE	923	6.239	26.400	14.153	1.00	16.57
ATOM	121	CB	ILE	923	6.775	25.894	12.783	1.00	17.14
ATOM	122	CG2	ILE	923	7.410	27.054	12.011	1.00	14.50
ATOM	123	CG1	ILE	923	7.782	24.756	12.971	1.00	13.94
ATOM	124	CD1	ILE	923	8.047	23.979	11.704	1.00	13.69
ATOM	125	C	ILE	923	5.530	25.272	14.912	1.00	18.05
ATOM	126	O	ILE	923	4.846	24.439	14.319	1.00	18.28
ATOM	127	N	LYS	924	5.710	25.264	16.230	1.00	19.42
ATOM	128	H	LYS	924	6.267	25.957	16.639	1.00	10.00
ATOM	129	CA	LYS	924	5.100	24.287	17.130	1.00	20.32
ATOM	130	CB	LYS	924	3.612	24.605	17.319	1.00	23.93
ATOM	131	CG	LYS	924	3.384	25.983	17.936	1.00	28.19
ATOM	132	CD	LYS	924	1.925	26.241	18.251	1.00	34.63
ATOM	133	CE	LYS	924	1.746	27.585	18.949	1.00	38.73
ATOM	134	NZ	LYS	924	0.313	27.854	19.284	1.00	44.05
ATOM	135	HZ1	LYS	924	-0.035	27.120	19.931	1.00	10.00
ATOM	136	HZ2	LYS	924	-0.257	27.845	18.412	1.00	10.00
ATOM	137	HZ3	LYS	924	0.227	28.791	19.730	1.00	10.00
ATOM	138	C	LYS	924	5.326	22.816	16.781	1.00	18.87
ATOM	139	O	LYS	924	4.485	21.956	17.054	1.00	17.67
ATOM	140	N	MET	925	6.497	22.530	16.218	1.00	18.00
ATOM	141	H	MET	925	7.115	23.269	16.046	1.00	10.00
ATOM	142	CA	MET	925	6.878	21.168	15.848	1.00	16.42
ATOM	143	CB	MET	925	7.227	21.093	14.358	1.00	15.95
ATOM	144	CG	MET	925	6.033	21.236	13.427	1.00	17.12
ATOM	145	SD	MET	925	4.947	19.804	13.475	1.00	18.04
ATOM	146	CE	MET	925	3.547	20.453	14.365	1.00	22.39
ATOM	147	C	MET	925	8.080	20.731	16.680	1.00	15.49
ATOM	148	O	MET	925	8.801	19.813	16.300	1.00	15.05
ATOM	149	N	ASP	926	8.265	21.364	17.834	1.00	15.42
ATOM	150	H	ASP	926	7.628	22.064	18.079	1.00	10.00
ATOM	151	CA	ASP	926	9.387	21.053	18.716	1.00	16.61
ATOM	152	CB	ASP	926	9.600	22.158	19.768	1.00	18.40
ATOM	153	CG	ASP	926	8.342	22.492	20.550	1.00	22.45
ATOM	154	OD1	ASP	926	8.366	22.335	21.785	1.00	29.54
ATOM	155	OD2	ASP	926	7.343	22.944	19.951	1.00	23.78
ATOM	156	C	ASP	926	9.349	19.660	19.351	1.00	17.08
ATOM	157	O	ASP	926	10.315	19.229	19.987	1.00	17.30
ATOM	158	N	ARG	927	8.260	18.932	19.126	1.00	15.97
ATOM	159	H	ARG	927	7.519	19.305	18.614	1.00	10.00
ATOM	160	CA	ARG	927	8.140	17.576	19.645	1.00	14.49
ATOM	161	CB	ARG	927	6.691	17.083	19.528	1.00	13.98
ATOM	162	CG	ARG	927	6.249	16.700	18.111	1.00	13.40
ATOM	163	CD	ARG	927	4.780	16.271	18.058	1.00	15.27
ATOM	164	NE	ARG	927	3.884	17.421	18.158	1.00	19.72
ATOM	165	HE	ARG	927	4.290	18.300	18.276	1.00	10.00
ATOM	166	CZ	ARG	927	2.557	17.356	18.111	1.00	18.27
ATOM	167	NH1	ARG	927	1.946	16.188	17.973	1.00	14.35
ATOM	168	HH11	ARG	927	2.481	15.348	17.913	1.00	0.00
ATOM	169	HH12	ARG	927	0.946	16.148	17.941	1.00	0.00
ATOM	170	NH2	ARG	927	1.840	18.472	18.178	1.00	17.89
ATOM	171	HH21	ARG	927	2.306	19.353	18.264	1.00	0.00
ATOM	172	HH22	ARG	927	0.840	18.431	18.143	1.00	0.00

ATOM	173	C	ARG	927	9.072	16.682	18.822	1.00	14.29
ATOM	174	O	ARG	927	9.361	15.547	19.204	1.00	14.14
ATOM	175	N	TYR	928	9.536	17.219	17.691	1.00	14.12
ATOM	176	H	TYR	928	9.268	18.120	17.426	1.00	10.00
ATOM	177	CA	TYR	928	10.431	16.522	16.770	1.00	13.74
ATOM	178	CB	TYR	928	9.938	16.706	15.332	1.00	11.65
ATOM	179	CG	TYR	928	8.610	16.029	15.057	1.00	12.93
ATOM	180	CD1	TYR	928	7.501	16.774	14.659	1.00	10.58
ATOM	181	CE1	TYR	928	6.288	16.171	14.386	1.00	10.49
ATOM	182	CD2	TYR	928	8.463	14.645	15.181	1.00	12.08
ATOM	183	CE2	TYR	928	7.236	14.020	14.907	1.00	12.39
ATOM	184	CZ	TYR	928	6.159	14.797	14.509	1.00	10.20
ATOM	185	OH	TYR	928	4.945	14.241	14.204	1.00	11.30
ATOM	186	HH	TYR	928	4.612	15.000	13.786	1.00	10.00
ATOM	187	C	TYR	928	11.904	16.934	16.857	1.00	14.27
ATOM	188	O	TYR	928	12.714	16.495	16.043	1.00	15.38
ATOM	189	N	LYS	929	12.255	17.749	17.849	1.00	16.73
ATOM	190	H	LYS	929	11.563	18.045	18.482	1.00	10.00
ATOM	191	CA	LYS	929	13.633	18.208	18.031	1.00	17.93
ATOM	192	CB	LYS	929	13.804	18.863	19.402	1.00	19.71
ATOM	193	CG	LYS	929	13.604	20.361	19.404	1.00	27.71
ATOM	194	CD	LYS	929	13.928	20.967	20.765	1.00	32.21
ATOM	195	CE	LYS	929	14.009	22.491	20.690	1.00	32.72
ATOM	196	NZ	LYS	929	12.749	23.124	20.175	1.00	31.10
ATOM	197	HZ1	LYS	929	11.938	22.865	20.776	1.00	10.00
ATOM	198	HZ2	LYS	929	12.566	22.800	19.207	1.00	10.00
ATOM	199	HZ3	LYS	929	12.856	24.158	20.192	1.00	10.00
ATOM	200	C	LYS	929	14.653	17.084	17.892	1.00	19.40
ATOM	201	O	LYS	929	15.566	17.149	17.064	1.00	18.56
ATOM	202	N	ASP	930	14.460	16.040	18.691	1.00	20.41
ATOM	203	H	ASP	930	13.694	16.063	19.294	1.00	10.00
ATOM	204	CA	ASP	930	15.347	14.887	18.702	1.00	20.42
ATOM	205	CB	ASP	930	14.958	13.932	19.836	1.00	23.30
ATOM	206	CG	ASP	930	15.210	14.523	21.214	1.00	25.30
ATOM	207	OD1	ASP	930	14.734	13.926	22.201	1.00	27.38
ATOM	208	OD2	ASP	930	15.885	15.575	21.317	1.00	26.70
ATOM	209	C	ASP	930	15.390	14.144	17.374	1.00	17.86
ATOM	210	O	ASP	930	16.417	13.561	17.031	1.00	17.51
ATOM	211	N	ASN	931	14.282	14.171	16.635	1.00	14.65
ATOM	212	H	ASN	931	13.530	14.700	16.955	1.00	10.00
ATOM	213	CA	ASN	931	14.200	13.504	15.335	1.00	13.52
ATOM	214	CB	ASN	931	12.788	13.625	14.753	1.00	13.46
ATOM	215	CG	ASN	931	11.742	12.911	15.590	1.00	12.15
ATOM	216	OD1	ASN	931	11.114	11.957	15.138	1.00	15.43
ATOM	217	ND2	ASN	931	11.531	13.386	16.810	1.00	10.00
ATOM	218	HD21	ASN	931	12.030	14.158	17.132	1.00	0.00
ATOM	219	HD22	ASN	931	10.842	12.912	17.313	1.00	0.00
ATOM	220	C	ASN	931	15.220	14.106	14.363	1.00	14.47
ATOM	221	O	ASN	931	15.917	13.382	13.654	1.00	13.19
ATOM	222	N	PHE	932	15.319	15.431	14.350	1.00	14.74
ATOM	223	H	PHE	932	14.747	15.968	14.943	1.00	10.00
ATOM	224	CA	PHE	932	16.266	16.119	13.478	1.00	16.13
ATOM	225	CB	PHE	932	15.963	17.621	13.441	1.00	13.55
ATOM	226	CG	PHE	932	14.757	17.965	12.615	1.00	13.40
ATOM	227	CD1	PHE	932	13.489	17.942	13.174	1.00	12.04
ATOM	228	CD2	PHE	932	14.888	18.271	11.269	1.00	12.73
ATOM	229	CE1	PHE	932	12.370	18.212	12.408	1.00	11.44
ATOM	230	CE2	PHE	932	13.771	18.543	10.493	1.00	12.55
ATOM	231	CZ	PHE	932	12.511	18.512	11.066	1.00	12.81

ATOM	232	C	PHE	932	17.708	15.875	13.914	1.00	17.19
ATOM	233	O	PHE	932	18.544	15.470	13.109	1.00	18.79
ATOM	234	N	THR	933	17.968	16.059	15.205	1.00	18.93
ATOM	235	H	THR	933	17.242	16.348	15.800	1.00	10.00
ATOM	236	CA	THR	933	19.293	15.878	15.779	1.00	19.46
ATOM	237	CB	THR	933	19.269	16.179	17.294	1.00	20.06
ATOM	238	OG1	THR	933	18.741	17.494	17.509	1.00	23.65
ATOM	239	HG1	THR	933	19.273	18.153	17.052	1.00	10.00
ATOM	240	CG2	THR	933	20.666	16.100	17.891	1.00	22.07
ATOM	241	C	THR	933	19.860	14.476	15.553	1.00	19.94
ATOM	242	O	THR	933	20.987	14.323	15.078	1.00	20.57
ATOM	243	N	ALA	934	19.068	13.457	15.871	1.00	19.21
ATOM	244	H	ALA	934	18.162	13.633	16.194	1.00	10.00
ATOM	245	CA	ALA	934	19.504	12.073	15.722	1.00	17.60
ATOM	246	CB	ALA	934	18.426	11.120	16.217	1.00	15.43
ATOM	247	C	ALA	934	19.910	11.717	14.301	1.00	16.03
ATOM	248	O	ALA	934	20.743	10.837	14.100	1.00	17.80
ATOM	249	N	ALA	935	19.335	12.408	13.321	1.00	16.14
ATOM	250	H	ALA	935	18.692	13.119	13.520	1.00	10.00
ATOM	251	CA	ALA	935	19.645	12.144	11.920	1.00	14.43
ATOM	252	CB	ALA	935	18.381	12.177	11.094	1.00	14.61
ATOM	253	C	ALA	935	20.692	13.091	11.337	1.00	13.86
ATOM	254	O	ALA	935	20.924	13.101	10.127	1.00	15.04
ATOM	255	N	GLY	936	21.328	13.878	12.197	1.00	13.66
ATOM	256	H	GLY	936	21.122	13.837	13.153	1.00	10.00
ATOM	257	CA	GLY	936	22.354	14.796	11.736	1.00	16.69
ATOM	258	C	GLY	936	21.878	16.152	11.247	1.00	17.44
ATOM	259	O	GLY	936	22.689	16.954	10.787	1.00	18.41
ATOM	260	N	TYR	937	20.573	16.412	11.312	1.00	19.20
ATOM	261	H	TYR	937	19.964	15.731	11.666	1.00	10.00
ATOM	262	CA	TYR	937	20.034	17.705	10.883	1.00	17.59
ATOM	263	CB	TYR	937	18.572	17.581	10.469	1.00	14.83
ATOM	264	CG	TYR	937	18.399	16.809	9.194	1.00	15.36
ATOM	265	CD1	TYR	937	18.107	15.452	9.220	1.00	15.74
ATOM	266	CE1	TYR	937	17.966	14.726	8.048	1.00	16.34
ATOM	267	CD2	TYR	937	18.551	17.427	7.960	1.00	14.36
ATOM	268	CE2	TYR	937	18.412	16.710	6.779	1.00	17.13
ATOM	269	CZ	TYR	937	18.120	15.360	6.831	1.00	15.80
ATOM	270	OH	TYR	937	17.962	14.662	5.665	1.00	17.66
ATOM	271	HH	TYR	937	18.108	15.268	4.961	1.00	10.00
ATOM	272	C	TYR	937	20.171	18.685	12.030	1.00	17.88
ATOM	273	O	TYR	937	19.223	18.919	12.783	1.00	18.66
ATOM	274	N	THR	938	21.369	19.242	12.163	1.00	19.08
ATOM	275	H	THR	938	22.094	18.997	11.546	1.00	10.00
ATOM	276	CA	THR	938	21.668	20.182	13.230	1.00	20.43
ATOM	277	CB	THR	938	23.015	19.845	13.891	1.00	20.49
ATOM	278	OG1	THR	938	24.024	19.724	12.880	1.00	23.01
ATOM	279	HG1	THR	938	24.865	19.550	13.311	1.00	10.00
ATOM	280	CG2	THR	938	22.912	18.538	14.666	1.00	21.47
ATOM	281	C	THR	938	21.666	21.643	12.816	1.00	20.47
ATOM	282	O	THR	938	21.740	22.524	13.672	1.00	22.23
ATOM	283	N	THR	939	21.610	21.906	11.515	1.00	19.49
ATOM	284	H	THR	939	21.592	21.174	10.869	1.00	10.00
ATOM	285	CA	THR	939	21.588	23.282	11.027	1.00	22.03
ATOM	286	CB	THR	939	22.919	23.679	10.306	1.00	21.88
ATOM	287	OG1	THR	939	23.138	22.832	9.170	1.00	23.38
ATOM	288	HG1	THR	939	22.419	22.887	8.550	1.00	10.00
ATOM	289	CG2	THR	939	24.112	23.567	11.255	1.00	22.58
ATOM	290	C	THR	939	20.414	23.478	10.072	1.00	22.29

ATOM	291	O	THR	939	19.983	22.533	9.403	1.00	23.11
ATOM	292	N	LEU	940	19.892	24.699	10.016	1.00	20.97
ATOM	293	H	LEU	940	20.257	25.408	10.589	1.00	10.00
ATOM	294	CA	LEU	940	18.778	25.008	9.127	1.00	21.45
ATOM	295	CB	LEU	940	18.231	26.411	9.396	1.00	21.73
ATOM	296	CG	LEU	940	17.245	26.592	10.553	1.00	20.62
ATOM	297	CD1	LEU	940	16.837	28.057	10.656	1.00	19.11
ATOM	298	CD2	LEU	940	16.015	25.721	10.330	1.00	22.59
ATOM	299	C	LEU	940	19.172	24.869	7.656	1.00	21.08
ATOM	300	O	LEU	940	18.313	24.702	6.796	1.00	19.68
ATOM	301	N	GLU	941	20.470	24.926	7.371	1.00	22.89
ATOM	302	H	GLU	941	21.105	25.087	8.093	1.00	10.00
ATOM	303	CA	GLU	941	20.956	24.782	5.998	1.00	24.55
ATOM	304	CB	GLU	941	22.467	25.046	5.931	1.00	29.29
ATOM	305	CG	GLU	941	22.890	26.397	6.509	1.00	39.55
ATOM	306	CD	GLU	941	24.371	26.712	6.317	1.00	45.17
ATOM	307	OE1	GLU	941	24.698	27.902	6.088	1.00	47.24
ATOM	308	OE2	GLU	941	25.205	25.780	6.406	1.00	46.83
ATOM	309	C	GLU	941	20.640	23.367	5.506	1.00	23.01
ATOM	310	O	GLU	941	20.234	23.169	4.361	1.00	23.35
ATOM	311	N	ALA	942	20.799	22.393	6.397	1.00	20.90
ATOM	312	H	ALA	942	21.090	22.606	7.301	1.00	10.00
ATOM	313	CA	ALA	942	20.531	20.998	6.080	1.00	19.22
ATOM	314	CB	ALA	942	21.183	20.096	7.115	1.00	18.33
ATOM	315	C	ALA	942	19.026	20.740	6.016	1.00	19.76
ATOM	316	O	ALA	942	18.547	20.018	5.140	1.00	19.96
ATOM	317	N	VAL	943	18.283	21.357	6.934	1.00	18.87
ATOM	318	H	VAL	943	18.728	21.923	7.601	1.00	10.00
ATOM	319	CA	VAL	943	16.831	21.203	6.991	1.00	16.47
ATOM	320	CB	VAL	943	16.233	21.956	8.210	1.00	14.10
ATOM	321	CG1	VAL	943	14.724	21.972	8.130	1.00	13.60
ATOM	322	CG2	VAL	943	16.665	21.297	9.509	1.00	13.04
ATOM	323	C	VAL	943	16.165	21.715	5.718	1.00	17.56
ATOM	324	O	VAL	943	15.303	21.054	5.142	1.00	18.81
ATOM	325	N	VAL	944	16.611	22.874	5.256	1.00	18.22
ATOM	326	H	VAL	944	17.343	23.321	5.728	1.00	10.00
ATOM	327	CA	VAL	944	16.056	23.511	4.076	1.00	19.55
ATOM	328	CB	VAL	944	16.627	24.961	3.944	1.00	20.45
ATOM	329	CG1	VAL	944	17.814	25.028	2.986	1.00	22.10
ATOM	330	CG2	VAL	944	15.537	25.919	3.577	1.00	20.47
ATOM	331	C	VAL	944	16.193	22.694	2.779	1.00	21.31
ATOM	332	O	VAL	944	15.579	23.019	1.760	1.00	21.60
ATOM	333	N	HIS	945	16.983	21.626	2.815	1.00	22.23
ATOM	334	H	HIS	945	17.461	21.383	3.636	1.00	10.00
ATOM	335	CA	HIS	945	17.148	20.776	1.638	1.00	24.47
ATOM	336	CB	HIS	945	18.627	20.625	1.270	1.00	29.45
ATOM	337	CG	HIS	945	19.258	21.897	0.787	1.00	35.62
ATOM	338	CD2	HIS	945	20.466	22.447	1.041	1.00	37.57
ATOM	339	ND1	HIS	945	18.596	22.785	-0.036	1.00	39.22
ATOM	340	HD1	HIS	945	17.684	22.691	-0.399	1.00	10.00
ATOM	341	CE1	HIS	945	19.370	23.830	-0.263	1.00	39.31
ATOM	342	NE2	HIS	945	20.509	23.653	0.379	1.00	40.84
ATOM	343	HE2	HIS	945	21.255	24.292	0.385	1.00	10.00
ATOM	344	C	HIS	945	16.493	19.405	1.813	1.00	23.50
ATOM	345	O	HIS	945	16.784	18.468	1.068	1.00	25.20
ATOM	346	N	MET	946	15.606	19.294	2.794	1.00	19.74
ATOM	347	H	MET	946	15.384	20.079	3.341	1.00	10.00
ATOM	348	CA	MET	946	14.908	18.045	3.049	1.00	18.63
ATOM	349	CB	MET	946	14.359	18.014	4.473	1.00	17.73

ATOM	350	CG	MET	946	15.406	17.951	5.545	1.00	19.84
ATOM	351	SD	MET	946	14.648	17.941	7.167	1.00	21.77
ATOM	352	CE	MET	946	14.493	16.171	7.450	1.00	18.09
ATOM	353	C	MET	946	13.745	17.867	2.089	1.00	18.12
ATOM	354	O	MET	946	13.234	18.834	1.523	1.00	18.77
ATOM	355	N	SER	947	13.327	16.617	1.927	1.00	17.33
ATOM	356	H	SER	947	13.780	15.896	2.405	1.00	10.00
ATOM	357	CA	SER	947	12.198	16.267	1.078	1.00	15.98
ATOM	358	CB	SER	947	12.641	15.297	-0.022	1.00	16.41
ATOM	359	OG	SER	947	13.144	14.084	0.523	1.00	14.22
ATOM	360	HG	SER	947	12.472	13.637	1.029	1.00	10.00
ATOM	361	C	SER	947	11.193	15.585	1.999	1.00	15.60
ATOM	362	O	SER	947	11.484	15.367	3.182	1.00	15.48
ATOM	363	N	GLN	948	10.033	15.221	1.461	1.00	15.65
ATOM	364	H	GLN	948	9.852	15.411	0.521	1.00	10.00
ATOM	365	CA	GLN	948	9.011	14.543	2.251	1.00	16.60
ATOM	366	CB	GLN	948	7.764	14.268	1.395	1.00	21.94
ATOM	367	CG	GLN	948	6.567	13.726	2.190	1.00	27.32
ATOM	368	CD	GLN	948	5.542	13.007	1.323	1.00	31.25
ATOM	369	OE1	GLN	948	5.303	13.382	0.175	1.00	34.83
ATOM	370	NE2	GLN	948	4.941	11.958	1.868	1.00	32.46
ATOM	371	HE21	GLN	948	5.180	11.719	2.787	1.00	10.00
ATOM	372	HE22	GLN	948	4.291	11.486	1.312	1.00	10.00
ATOM	373	C	GLN	948	9.583	13.216	2.768	1.00	15.21
ATOM	374	O	GLN	948	9.406	12.864	3.936	1.00	12.66
ATOM	375	N	ASP	949	10.304	12.511	1.900	1.00	13.65
ATOM	376	H	ASP	949	10.406	12.861	0.994	1.00	10.00
ATOM	377	CA	ASP	949	10.901	11.223	2.251	1.00	14.81
ATOM	378	CB	ASP	949	11.553	10.574	1.022	1.00	15.09
ATOM	379	CG	ASP	949	12.016	9.153	1.286	1.00	17.60
ATOM	380	OD1	ASP	949	13.236	8.913	1.211	1.00	18.94
ATOM	381	OD2	ASP	949	11.167	8.276	1.564	1.00	19.58
ATOM	382	C	ASP	949	11.905	11.352	3.386	1.00	13.01
ATOM	383	O	ASP	949	12.030	10.449	4.205	1.00	14.17
ATOM	384	N	ASP	950	12.608	12.480	3.438	1.00	13.51
ATOM	385	H	ASP	950	12.479	13.181	2.775	1.00	10.00
ATOM	386	CA	ASP	950	13.580	12.717	4.501	1.00	13.81
ATOM	387	CB	ASP	950	14.390	13.990	4.234	1.00	14.10
ATOM	388	CG	ASP	950	15.410	13.813	3.120	1.00	14.34
ATOM	389	OD1	ASP	950	15.924	12.691	2.941	1.00	16.13
ATOM	390	OD2	ASP	950	15.706	14.798	2.422	1.00	16.24
ATOM	391	C	ASP	950	12.861	12.830	5.838	1.00	13.61
ATOM	392	O	ASP	950	13.312	12.272	6.834	1.00	14.29
ATOM	393	N	LEU	951	11.721	13.521	5.843	1.00	12.11
ATOM	394	H	LEU	951	11.404	13.921	5.005	1.00	10.00
ATOM	395	CA	LEU	951	10.923	13.702	7.054	1.00	12.52
ATOM	396	CB	LEU	951	9.802	14.722	6.815	1.00	11.55
ATOM	397	CG	LEU	951	10.264	16.155	6.522	1.00	13.10
ATOM	398	CD1	LEU	951	9.085	17.016	6.102	1.00	11.51
ATOM	399	CD2	LEU	951	10.956	16.743	7.740	1.00	10.00
ATOM	400	C	LEU	951	10.345	12.376	7.551	1.00	11.50
ATOM	401	O	LEU	951	10.320	12.122	8.755	1.00	12.61
ATOM	402	N	ALA	952	9.887	11.532	6.628	1.00	11.25
ATOM	403	H	ALA	952	9.920	11.800	5.686	1.00	10.00
ATOM	404	CA	ALA	952	9.341	10.229	6.991	1.00	11.84
ATOM	405	CB	ALA	952	8.741	9.544	5.767	1.00	10.00
ATOM	406	C	ALA	952	10.455	9.363	7.598	1.00	13.43
ATOM	407	O	ALA	952	10.278	8.750	8.654	1.00	14.76
ATOM	408	N	ARG	953	11.614	9.361	6.944	1.00	12.41

ATOM	409	H	ARG	953	11.707	9.895	6.128	1.00	10.00
ATOM	410	CA	ARG	953	12.776	8.591	7.390	1.00	15.00
ATOM	411	CB	ARG	953	13.921	8.763	6.381	1.00	15.69
ATOM	412	CG	ARG	953	15.223	8.077	6.762	1.00	22.48
ATOM	413	CD	ARG	953	16.299	8.246	5.686	1.00	24.87
ATOM	414	NE	ARG	953	16.604	9.651	5.409	1.00	25.58
ATOM	415	HE	ARG	953	16.130	10.085	4.673	1.00	10.00
ATOM	416	CZ	ARG	953	17.475	10.387	6.095	1.00	25.21
ATOM	417	NH1	ARG	953	18.146	9.863	7.115	1.00	24.46
ATOM	418	HH11	ARG	953	17.997	8.911	7.378	1.00	0.00
ATOM	419	HH12	ARG	953	18.796	10.425	7.626	1.00	0.00
ATOM	420	NH2	ARG	953	17.668	11.656	5.765	1.00	22.85
ATOM	421	HH21	ARG	953	17.157	12.052	5.000	1.00	0.00
ATOM	422	HH22	ARG	953	18.321	12.214	6.276	1.00	0.00
ATOM	423	C	ARG	953	13.225	9.028	8.786	1.00	15.66
ATOM	424	O	ARG	953	13.771	8.237	9.551	1.00	16.37
ATOM	425	N	ILE	954	12.947	10.285	9.115	1.00	16.54
ATOM	426	H	ILE	954	12.499	10.847	8.451	1.00	10.00
ATOM	427	CA	ILE	954	13.303	10.891	10.396	1.00	15.90
ATOM	428	CB	ILE	954	13.518	12.437	10.194	1.00	18.15
ATOM	429	CG2	ILE	954	12.974	13.269	11.338	1.00	20.90
ATOM	430	CG1	ILE	954	14.997	12.722	9.975	1.00	18.92
ATOM	431	CD1	ILE	954	15.609	11.919	8.863	1.00	23.36
ATOM	432	C	ILE	954	12.305	10.573	11.519	1.00	14.18
ATOM	433	O	ILE	954	12.613	10.743	12.700	1.00	12.65
ATOM	434	N	GLY	955	11.135	10.065	11.153	1.00	13.39
ATOM	435	H	GLY	955	10.942	9.909	10.205	1.00	10.00
ATOM	436	CA	GLY	955	10.136	9.736	12.154	1.00	14.64
ATOM	437	C	GLY	955	8.901	10.616	12.114	1.00	14.46
ATOM	438	O	GLY	955	7.993	10.462	12.930	1.00	17.04
ATOM	439	N	ILE	956	8.862	11.558	11.181	1.00	14.53
ATOM	440	H	ILE	956	9.610	11.653	10.555	1.00	10.00
ATOM	441	CA	ILE	956	7.710	12.441	11.055	1.00	15.41
ATOM	442	CB	ILE	956	8.142	13.883	10.676	1.00	14.68
ATOM	443	CG2	ILE	956	6.925	14.817	10.645	1.00	13.49
ATOM	444	CG1	ILE	956	9.186	14.384	11.683	1.00	14.55
ATOM	445	CD1	ILE	956	9.793	15.712	11.340	1.00	14.38
ATOM	446	C	ILE	956	6.819	11.829	9.977	1.00	14.93
ATOM	447	O	ILE	956	6.954	12.128	8.789	1.00	15.07
ATOM	448	N	THR	957	5.937	10.934	10.408	1.00	16.86
ATOM	449	H	THR	957	5.929	10.728	11.366	1.00	10.00
ATOM	450	CA	THR	957	5.031	10.231	9.506	1.00	17.30
ATOM	451	CB	THR	957	4.822	8.779	9.960	1.00	17.04
ATOM	452	OG1	THR	957	4.334	8.770	11.308	1.00	19.91
ATOM	453	HG1	THR	957	4.987	9.124	11.917	1.00	10.00
ATOM	454	CG2	THR	957	6.134	8.005	9.887	1.00	18.15
ATOM	455	C	THR	957	3.671	10.890	9.336	1.00	16.02
ATOM	456	O	THR	957	2.987	10.638	8.349	1.00	16.25
ATOM	457	N	ALA	958	3.269	11.710	10.304	1.00	16.24
ATOM	458	H	ALA	958	3.857	11.836	11.074	1.00	10.00
ATOM	459	CA	ALA	958	1.982	12.406	10.223	1.00	15.66
ATOM	460	CB	ALA	958	1.680	13.140	11.536	1.00	10.94
ATOM	461	C	ALA	958	2.063	13.399	9.067	1.00	14.30
ATOM	462	O	ALA	958	2.830	14.365	9.128	1.00	14.94
ATOM	463	N	ILE	959	1.282	13.156	8.019	1.00	14.35
ATOM	464	H	ILE	959	0.709	12.360	8.024	1.00	10.00
ATOM	465	CA	ILE	959	1.286	14.017	6.838	1.00	16.17
ATOM	466	CB	ILE	959	0.286	13.521	5.769	1.00	17.13
ATOM	467	CG2	ILE	959	0.351	14.410	4.532	1.00	17.99

	ATOM	468	CG1	ILE	959	0.622	12.078	5.371	1.00	20.81
	ATOM	469	CD1	ILE	959	-0.355	11.449	4.383	1.00	21.87
	ATOM	470	C	ILE	959	1.042	15.493	7.163	1.00	16.99
	ATOM	471	O	ILE	959	1.661	16.374	6.564	1.00	18.92
	ATOM	472	N	THR	960	0.179	15.765	8.137	1.00	16.32
	ATOM	473	H	THR	960	-0.303	15.030	8.563	1.00	10.00
	ATOM	474	CA	THR	960	-0.103	17.142	8.534	1.00	16.86
	ATOM	475	CB	THR	960	-1.165	17.190	9.660	1.00	18.03
	ATOM	476	OG1	THR	960	-2.384	16.597	9.197	1.00	18.57
	ATOM	477	HG1	THR	960	-3.070	16.619	9.863	1.00	10.00
	ATOM	478	CG2	THR	960	-1.438	18.618	10.082	1.00	18.31
	ATOM	479	C	THR	960	1.180	17.824	9.026	1.00	16.90
	ATOM	480	O	THR	960	1.465	18.974	8.683	1.00	16.92
	ATOM	481	N	HIS	961	1.955	17.105	9.830	1.00	16.15
	ATOM	482	H	HIS	961	1.713	16.175	10.007	1.00	10.00
	ATOM	483	CA	HIS	961	3.197	17.646	10.364	1.00	15.68
	ATOM	484	CB	HIS	961	3.689	16.789	11.532	1.00	15.19
	ATOM	485	CG	HIS	961	2.758	16.807	12.706	1.00	14.80
	ATOM	486	CD2	HIS	961	1.715	17.624	12.994	1.00	14.08
	ATOM	487	ND1	HIS	961	2.834	15.909	13.747	1.00	14.93
	ATOM	488	HD1	HIS	961	3.301	15.070	13.827	1.00	10.00
	ATOM	489	CE1	HIS	961	1.886	16.172	14.625	1.00	16.16
	ATOM	490	NE2	HIS	961	1.191	17.207	14.192	1.00	14.72
	ATOM	491	HE2	HIS	961	0.422	17.616	14.676	1.00	10.00
	ATOM	492	C	HIS	961	4.248	17.788	9.268	1.00	15.94
	ATOM	493	O	HIS	961	5.020	18.750	9.272	1.00	14.39
	ATOM	494	N	GLN	962	4.242	16.853	8.316	1.00	16.90
	ATOM	495	H	GLN	962	3.608	16.106	8.356	1.00	10.00
	ATOM	496	CA	GLN	962	5.167	16.895	7.185	1.00	16.19
	ATOM	497	CB	GLN	962	4.959	15.694	6.254	1.00	15.77
	ATOM	498	CG	GLN	962	5.490	14.367	6.780	1.00	16.87
	ATOM	499	CD	GLN	962	5.269	13.218	5.802	1.00	17.54
	ATOM	500	OE1	GLN	962	4.677	13.396	4.737	1.00	18.11
	ATOM	501	NE2	GLN	962	5.743	12.036	6.163	1.00	17.65
	ATOM	502	HE21	GLN	962	6.214	11.964	7.022	1.00	10.00
	ATOM	503	HE22	GLN	962	5.602	11.288	5.549	1.00	10.00
	ATOM	504	C	GLN	962	4.903	18.173	6.401	1.00	15.74
	ATOM	505	O	GLN	962	5.832	18.899	6.059	1.00	15.23
	ATOM	506	N	ASN	963	3.625	18.463	6.168	1.00	15.23
	ATOM	507	H	ASN	963	2.932	17.852	6.489	1.00	10.00
	ATOM	508	CA	ASN	963	3.232	19.649	5.415	1.00	16.00
	ATOM	509	CB	ASN	963	1.767	19.562	4.981	1.00	16.74
	ATOM	510	CG	ASN	963	1.567	18.605	3.821	1.00	20.62
	ATOM	511	OD1	ASN	963	2.379	18.560	2.895	1.00	24.90
	ATOM	512	ND2	ASN	963	0.510	17.811	3.879	1.00	22.06
	ATOM	513	HD21	ASN	963	-0.085	17.835	4.647	1.00	0.00
	ATOM	514	HD22	ASN	963	0.384	17.209	3.108	1.00	0.00
	ATOM	515	C	ASN	963	3.513	20.955	6.129	1.00	14.80
	ATOM	516	O	ASN	963	3.907	21.925	5.495	1.00	13.94
	ATOM	517	N	LYS	964	3.344	20.978	7.448	1.00	15.34
	ATOM	518	H	LYS	964	3.023	20.172	7.901	1.00	10.00
	ATOM	519	CA	LYS	964	3.618	22.191	8.208	1.00	16.38
	ATOM	520	CB	LYS	964	3.254	22.013	9.686	1.00	16.89
	ATOM	521	CG	LYS	964	3.415	23.289	10.501	1.00	20.93
	ATOM	522	CD	LYS	964	2.888	23.125	11.911	1.00	28.63
	ATOM	523	CE	LYS	964	1.917	24.244	12.280	1.00	31.73
	ATOM	524	NZ	LYS	964	2.519	25.607	12.214	1.00	33.69
	ATOM	525	HZ1	LYS	964	2.843	25.813	11.248	1.00	10.00
	ATOM	526	HZ2	LYS	964	3.314	25.670	12.878	1.00	10.00

	ATOM	527	HZ3	LYS	964	1.798	26.313	12.479	1.00	10.00
	ATOM	528	C	LYS	964	5.103	22.514	8.078	1.00	16.37
	ATOM	529	O	LYS	964	5.487	23.662	7.844	1.00	17.92
	ATOM	530	N	ILE	965	5.930	21.481	8.199	1.00	16.65
	ATOM	531	H	ILE	965	5.562	20.586	8.353	1.00	10.00
	ATOM	532	CA	ILE	965	7.374	21.629	8.099	1.00	14.00
	ATOM	533	CB	ILE	965	8.088	20.346	8.560	1.00	12.06
	ATOM	534	CG2	ILE	965	9.560	20.369	8.151	1.00	11.41
	ATOM	535	CG1	ILE	965	7.934	20.202	10.076	1.00	10.00
	ATOM	536	CD1	ILE	965	8.302	18.849	10.602	1.00	10.00
	ATOM	537	C	ILE	965	7.800	22.012	6.687	1.00	13.95
	ATOM	538	O	ILE	965	8.574	22.951	6.506	1.00	15.38
	ATOM	539	N	LEU	966	7.274	21.310	5.691	1.00	13.46
	ATOM	540	H	LEU	966	6.650	20.584	5.886	1.00	10.00
	ATOM	541	CA	LEU	966	7.609	21.605	4.304	1.00	15.13
	ATOM	542	CB	LEU	966	7.023	20.542	3.372	1.00	15.28
	ATOM	543	CG	LEU	966	7.738	19.187	3.463	1.00	14.45
	ATOM	544	CD1	LEU	966	6.910	18.105	2.791	1.00	14.55
	ATOM	545	CD2	LEU	966	9.123	19.285	2.841	1.00	12.40
	ATOM	546	C	LEU	966	7.173	23.015	3.886	1.00	15.31
	ATOM	547	O	LEU	966	7.909	23.712	3.181	1.00	13.98
	ATOM	548	N	SER	967	6.000	23.446	4.345	1.00	15.58
	ATOM	549	H	SER	967	5.435	22.863	4.889	1.00	10.00
	ATOM	550	CA	SER	967	5.506	24.780	4.029	1.00	15.48
	ATOM	551	CB	SER	967	4.073	24.974	4.537	1.00	16.99
	ATOM	552	OG	SER	967	3.142	24.292	3.715	1.00	19.67
	ATOM	553	HG	SER	967	2.272	24.346	4.118	1.00	10.00
	ATOM	554	C	SER	967	6.427	25.824	4.643	1.00	16.18
	ATOM	555	O	SER	967	6.725	26.839	4.017	1.00	16.88
	ATOM	556	N	SER	968	6.892	25.554	5.859	1.00	16.46
	ATOM	557	H	SER	968	6.624	24.721	6.303	1.00	10.00
	ATOM	558	CA	SER	968	7.789	26.459	6.570	1.00	17.49
	ATOM	559	CB	SER	968	8.024	25.947	7.995	1.00	17.79
	ATOM	560	OG	SER	968	8.766	26.871	8.774	1.00	16.89
	ATOM	561	HG	SER	968	8.248	27.676	8.828	1.00	10.00
	ATOM	562	C	SER	968	9.114	26.562	5.815	1.00	18.08
	ATOM	563	O	SER	968	9.661	27.654	5.650	1.00	18.78
	ATOM	564	N	VAL	969	9.621	25.418	5.360	1.00	17.77
	ATOM	565	H	VAL	969	9.145	24.581	5.543	1.00	10.00
	ATOM	566	CA	VAL	969	10.863	25.362	4.598	1.00	16.32
	ATOM	567	CB	VAL	969	11.201	23.892	4.189	1.00	15.01
	ATOM	568	CG1	VAL	969	12.232	23.855	3.065	1.00	14.69
	ATOM	569	CG2	VAL	969	11.729	23.123	5.390	1.00	12.25
	ATOM	570	C	VAL	969	10.720	26.256	3.360	1.00	16.88
	ATOM	571	O	VAL	969	11.597	27.072	3.071	1.00	18.31
	ATOM	572	N	GLN	970	9.590	26.140	2.666	1.00	17.34
	ATOM	573	H	GLN	970	8.915	25.497	2.971	1.00	10.00
	ATOM	574	CA	GLN	970	9.334	26.945	1.475	1.00	18.13
	ATOM	575	CB	GLN	970	7.977	26.588	0.859	1.00	17.14
	ATOM	576	CG	GLN	970	7.886	25.155	0.350	1.00	18.59
	ATOM	577	CD	GLN	970	6.520	24.809	-0.215	1.00	20.55
	ATOM	578	OE1	GLN	970	6.417	24.141	-1.237	1.00	23.07
	ATOM	579	NE2	GLN	970	5.466	25.251	0.453	1.00	21.52
	ATOM	580	HE21	GLN	970	5.590	25.769	1.273	1.00	10.00
	ATOM	581	HE22	GLN	970	4.591	25.025	0.074	1.00	10.00
	ATOM	582	C	GLN	970	9.391	28.442	1.792	1.00	19.00
	ATOM	583	O	GLN	970	9.998	29.216	1.047	1.00	19.59
	ATOM	584	N	ALA	971	8.793	28.832	2.915	1.00	18.42
	ATOM	585	H	ALA	971	8.345	28.158	3.469	1.00	10.00

ATOM	586	CA	ALA	971	8.772	30.227	3.351	1.00	18.65
ATOM	587	CB	ALA	971	7.797	30.397	4.510	1.00	16.83
ATOM	588	C	ALA	971	10.160	30.733	3.748	1.00	19.41
ATOM	589	O	ALA	971	10.512	31.884	3.477	1.00	20.57
ATOM	590	N	MET	972	10.945	29.880	4.399	1.00	20.08
ATOM	591	H	MET	972	10.613	28.976	4.582	1.00	10.00
ATOM	592	CA	MET	972	12.283	30.263	4.819	1.00	20.97
ATOM	593	CB	MET	972	12.876	29.248	5.788	1.00	23.16
ATOM	594	CG	MET	972	12.260	29.289	7.174	1.00	25.68
ATOM	595	SD	MET	972	13.275	28.393	8.360	1.00	29.72
ATOM	596	CE	MET	972	13.134	26.696	7.733	1.00	28.75
ATOM	597	C	MET	972	13.199	30.431	3.620	1.00	23.17
ATOM	598	O	MET	972	14.072	31.297	3.624	1.00	22.87
ATOM	599	N	ARG	973	13.002	29.603	2.597	1.00	22.37
ATOM	600	H	ARG	973	12.308	28.911	2.661	1.00	10.00
ATOM	601	CA	ARG	973	13.803	29.690	1.382	1.00	22.73
ATOM	602	CB	ARG	973	13.513	28.507	0.458	1.00	19.60
ATOM	603	CG	ARG	973	14.116	27.209	0.929	1.00	18.92
ATOM	604	CD	ARG	973	13.681	26.058	0.056	1.00	19.57
ATOM	605	NE	ARG	973	13.960	26.318	-1.353	1.00	23.42
ATOM	606	HE	ARG	973	13.341	26.895	-1.845	1.00	10.00
ATOM	607	CZ	ARG	973	14.994	25.817	-2.020	1.00	22.59
ATOM	608	NH1	ARG	973	15.862	25.015	-1.414	1.00	23.50
ATOM	609	HH11	ARG	973	15.747	24.775	-0.452	1.00	0.00
ATOM	610	HH12	ARG	973	16.628	24.635	-1.935	1.00	0.00
ATOM	611	NH2	ARG	973	15.167	26.135	-3.295	1.00	22.09
ATOM	612	HH21	ARG	973	14.515	26.744	-3.751	1.00	0.00
ATOM	613	HH22	ARG	973	15.933	25.750	-3.811	1.00	0.00
ATOM	614	C	ARG	973	13.507	31.010	0.670	1.00	24.33
ATOM	615	O	ARG	973	14.432	31.730	0.282	1.00	22.76
ATOM	616	N	THR	974	12.220	31.332	0.527	1.00	26.94
ATOM	617	H	THR	974	11.515	30.729	0.852	1.00	10.00
ATOM	618	CA	THR	974	11.795	32.576	-0.113	1.00	31.31
ATOM	619	CB	THR	974	10.264	32.669	-0.193	1.00	30.78
ATOM	620	OG1	THR	974	9.750	31.491	-0.822	1.00	31.97
ATOM	621	HG1	THR	974	10.092	31.425	-1.714	1.00	10.00
ATOM	622	CG2	THR	974	9.842	33.884	-1.005	1.00	32.02
ATOM	623	C	THR	974	12.325	33.783	0.671	1.00	35.46
ATOM	624	O	THR	974	12.720	34.792	0.081	1.00	36.00
ATOM	625	N	GLN	975	12.364	33.655	1.996	1.00	39.13
ATOM	626	H	GLN	975	12.031	32.830	2.414	1.00	10.00
ATOM	627	CA	GLN	975	12.858	34.715	2.869	1.00	42.57
ATOM	628	CB	GLN	975	12.586	34.359	4.334	1.00	45.95
ATOM	629	CG	GLN	975	12.793	35.519	5.300	1.00	51.94
ATOM	630	CD	GLN	975	12.891	35.080	6.751	1.00	54.33
ATOM	631	OE1	GLN	975	12.227	34.130	7.180	1.00	54.56
ATOM	632	NE2	GLN	975	13.728	35.773	7.517	1.00	55.74
ATOM	633	HE21	GLN	975	14.227	36.509	7.105	1.00	10.00
ATOM	634	HE22	GLN	975	13.811	35.543	8.462	1.00	10.00
ATOM	635	C	GLN	975	14.361	34.917	2.660	1.00	42.58
ATOM	636	O	GLN	975	14.862	36.043	2.706	1.00	43.50
ATOM	637	N	MET	976	15.072	33.818	2.424	1.00	42.59
ATOM	638	H	MET	976	14.606	32.954	2.412	1.00	10.00
ATOM	639	CA	MET	976	16.513	33.860	2.204	1.00	42.79
ATOM	640	CB	MET	976	17.129	32.480	2.416	1.00	41.79
ATOM	641	CG	MET	976	16.916	31.925	3.810	1.00	42.17
ATOM	642	SD	MET	976	17.517	33.018	5.102	1.00	44.85
ATOM	643	CE	MET	976	16.022	33.352	6.019	1.00	42.12
ATOM	644	C	MET	976	16.892	34.408	0.827	1.00	43.71

ATOM	645	O	MET	976	18.016	34.871	0.631	1.00	44.60
ATOM	646	N	GLN	977	15.975	34.327	-0.136	1.00	43.94
ATOM	647	H	GLN	977	15.105	33.913	0.057	1.00	10.00
ATOM	648	CA	GLN	977	16.235	34.859	-1.472	1.00	45.08
ATOM	649	CB	GLN	977	15.122	34.482	-2.449	1.00	46.20
ATOM	650	CG	GLN	977	15.070	33.029	-2.832	1.00	49.33
ATOM	651	CD	GLN	977	14.291	32.813	-4.112	1.00	52.53
ATOM	652	OE1	GLN	977	14.742	33.196	-5.195	1.00	52.71
ATOM	653	NE2	GLN	977	13.119	32.198	-4.000	1.00	53.68
ATOM	654	HE21	GLN	977	12.811	31.913	-3.116	1.00	10.00
ATOM	655	HE22	GLN	977	12.622	32.067	-4.836	1.00	10.00
ATOM	656	C	GLN	977	16.282	36.375	-1.370	1.00	45.82
ATOM	657	O	GLN	977	17.046	37.037	-2.070	1.00	45.80
ATOM	658	N	GLN	978	15.453	36.903	-0.475	1.00	47.73
ATOM	659	H	GLN	978	14.872	36.302	0.037	1.00	10.00
ATOM	660	CA	GLN	978	15.337	38.333	-0.224	1.00	48.95
ATOM	661	CB	GLN	978	14.007	38.606	0.482	1.00	50.40
ATOM	662	CG	GLN	978	12.800	38.062	-0.288	1.00	51.80
ATOM	663	CD	GLN	978	11.535	37.947	0.556	1.00	54.30
ATOM	664	OE1	GLN	978	10.439	37.770	0.021	1.00	53.56
ATOM	665	NE2	GLN	978	11.683	38.018	1.877	1.00	54.22
ATOM	666	HE21	GLN	978	12.559	38.123	2.294	1.00	10.00
ATOM	667	HE22	GLN	978	10.843	37.960	2.381	1.00	10.00
ATOM	668	C	GLN	978	16.511	38.860	0.604	1.00	49.37
ATOM	669	O	GLN	978	16.656	40.068	0.792	1.00	51.47
ATOM	670	N	MET	979	17.361	37.949	1.070	1.00	49.20
ATOM	671	H	MET	979	17.205	37.003	0.891	1.00	10.00
ATOM	672	CA	MET	979	18.532	38.309	1.863	1.00	49.49
ATOM	673	CB	MET	979	18.939	37.138	2.767	1.00	50.98
ATOM	674	CG	MET	979	19.641	37.532	4.064	1.00	52.98
ATOM	675	SD	MET	979	18.533	38.331	5.249	1.00	56.67
ATOM	676	CE	MET	979	17.092	37.225	5.208	1.00	55.34
ATOM	677	C	MET	979	19.702	38.672	0.941	1.00	48.66
ATOM	678	O	MET	979	20.842	38.788	1.392	1.00	48.62
ATOM	679	N	HIS	980	19.424	38.802	-0.356	1.00	46.50
ATOM	680	H	HIS	980	18.519	38.655	-0.702	1.00	10.00
ATOM	681	CA	HIS	980	20.454	39.155	-1.319	1.00	44.17
ATOM	682	CB	HIS	980	21.468	38.018	-1.487	1.00	39.73
ATOM	683	CG	HIS	980	20.883	36.743	-2.002	1.00	37.33
ATOM	684	CD2	HIS	980	20.507	35.611	-1.360	1.00	34.61
ATOM	685	ND1	HIS	980	20.667	36.504	-3.343	1.00	34.47
ATOM	686	HD1	HIS	980	20.768	37.181	-4.053	1.00	10.00
ATOM	687	CE1	HIS	980	20.192	35.286	-3.505	1.00	30.99
ATOM	688	NE2	HIS	980	20.087	34.722	-2.315	1.00	32.54
ATOM	689	HE2	HIS	980	19.880	33.801	-2.092	1.00	10.00
ATOM	690	C	HIS	980	19.910	39.599	-2.676	1.00	45.06
ATOM	691	O	HIS	980	20.126	38.942	-3.695	1.00	46.18
ATOM	692	N	GLY	981	19.171	40.703	-2.668	1.00	46.12
ATOM	693	H	GLY	981	18.996	41.168	-1.838	1.00	10.00
ATOM	694	CA	GLY	981	18.623	41.249	-3.900	1.00	46.12
ATOM	695	C	GLY	981	19.611	42.222	-4.526	1.00	46.17
ATOM	696	O	GLY	981	19.297	42.809	-5.583	1.00	45.93
ATOM	697	OT	GLY	981	20.710	42.404	-3.954	1.00	47.24
ATOM	698	OH2	TIP3	1	5.348	20.105	18.757	1.00	13.76
SOLV									
ATOM	699	OH2	TIP3	2	-1.607	18.597	5.643	1.00	10.67
SOLV									
ATOM	700	OH2	TIP3	3	11.575	6.309	4.064	1.00	22.41
SOLV									

```
ATOM    701  OH2 TIP3    4      -1.390  17.519  15.624  1.00 26.01
SOLV
ATOM    702  OH2 TIP3    5      10.169  12.836  -1.287  1.00 21.06
SOLV
ATOM    703  OH2 TIP3    6      22.240  14.887   7.379  1.00 23.85
SOLV
ATOM    704  OH2 TIP3    7       4.944  28.525   2.399  1.00 23.14
SOLV
ATOM    705  OH2 TIP3    8       9.075  16.556  -1.049  1.00 20.96
SOLV
ATOM    706  OH2 TIP3    9      20.954  26.572  12.360  1.00 24.79
SOLV
ATOM    707  OH2 TIP3   10      11.237  23.923  22.404  1.00 26.34
SOLV
ATOM    708  OH2 TIP3   12       2.003  11.736   0.983  1.00 32.92
SOLV
ATOM    709  OH2 TIP3   13      22.341  26.794   8.789  1.00 30.68
SOLV
ATOM    710  OH2 TIP3   14      20.229   8.193   9.431  1.00 32.28
SOLV
ATOM    711  OH2 TIP3   15       9.002   9.880  15.816  1.00 43.59
SOLV
ATOM    712  OH2 TIP3   16      -1.539  23.366  11.757  1.00 34.88
SOLV
ATOM    713  OH2 TIP3   18      17.039  21.091  18.454  1.00 24.01
SOLV
ATOM    714  OH2 TIP3   19       1.972  14.588   0.930  1.00 37.49
SOLV
ATOM    715  OH2 TIP3   20       2.846  13.524  16.999  1.00 38.34
SOLV
ATOM    716  OH2 TIP3   21      23.328  11.454  15.152  1.00 32.30
SOLV
ATOM    717  OH2 TIP3   22      20.264  11.491   3.039  1.00 37.02
SOLV
ATOM    718  OH2 TIP3   23      -1.623  13.755   9.215  1.00 24.03
SOLV
ATOM    719  OH2 TIP3   24       7.117   7.426   3.341  1.00 39.59
SOLV
ATOM    720  OH2 TIP3   25      12.549  12.170  21.440  1.00 39.58
SOLV
ATOM    721  OH2 TIP3   26      18.509  15.812   3.112  1.00 19.84
SOLV
ATOM    722  OH2 TIP3   27      -1.105  20.234  15.360  1.00 54.40
SOLV
ATOM    723  OH2 TIP3   28      13.308  28.872  -3.398  1.00 31.59
SOLV
ATOM    724  OH2 TIP3   29      20.963  24.454  14.891  1.00 29.47
SOLV
ATOM    725  OH2 TIP3   30      11.976  15.552  20.454  1.00 22.32
SOLV
ATOM    726  OH2 TIP3   31      15.358   9.999   2.685  1.00 25.88
SOLV
ATOM    727  OH2 TIP3   32       7.138  31.193  -0.433  1.00 34.94
SOLV
ATOM    728  OH2 TIP3   33      18.565  19.866  15.827  1.00 31.14
SOLV
ATOM    729  OH2 TIP3   35      10.191  11.998  19.068  1.00 36.35
SOLV
```

```
ATOM    730  OH2 TIP3   36      -1.668  14.793  13.334  1.00 32.50
SOLV
ATOM    731  OH2 TIP3   37      24.133  32.437  15.070  1.00 58.18
SOLV
ATOM    732  OH2 TIP3   38       8.142  29.716   8.516  1.00 38.02
SOLV
ATOM    733  OH2 TIP3   39      16.767  15.194  -0.104  1.00 22.13
SOLV
ATOM    734  OH2 TIP3   40      -0.649  26.330  15.687  1.00 40.04
SOLV
ATOM    735  OH2 TIP3   41      20.560  21.578  16.917  1.00 41.45
SOLV
ATOM    736  OH2 TIP3   42      24.262  18.640   8.488  1.00 46.62
SOLV
ATOM    737  OH2 TIP3   43       8.129   9.379   1.097  1.00 38.14
SOLV
ATOM    738  OH2 TIP3   44       4.176  17.058  22.386  1.00 43.87
SOLV
ATOM    739  OH2 TIP3   45       4.329  26.311   8.281  1.00 32.81
SOLV
ATOM    740  OH2 TIP3   46      19.760  18.168   3.397  1.00 26.43
SOLV
ATOM    741  OH2 TIP3   47       4.718  28.777   6.740  1.00 27.69
SOLV
ATOM    742  OH2 TIP3   48       7.659   9.629  -3.363  1.00 41.77
SOLV
ATOM    743  OH2 TIP3   49       2.827  14.980  21.015  1.00 38.06
SOLV
ATOM    744  OH2 TIP3   50       5.873  25.576  -4.108  1.00 58.77
SOLV
ATOM    745  OH2 TIP3   51      22.281  25.625  17.414  1.00 46.99
SOLV
ATOM    746  OH2 TIP3   53       8.311  41.084   2.721  1.00 41.45
SOLV
ATOM    747  OH2 TIP3   54      23.900  11.535  10.030  1.00 28.69
SOLV
ATOM    748  OH2 TIP3   55      23.435  27.423  11.489  1.00 40.75
SOLV
ATOM    749  OH2 TIP3   56      16.616  38.557  -4.772  1.00 41.46
SOLV
ATOM    750  OH2 TIP3   57      10.916   5.891  -0.754  1.00 26.82
SOLV
END
```

FULL CITATIONS FOR REFERENCES REFERRED TO IN THE SPECIFICATION

1. Schultz, J., Ponting, C.P., Hofmann, K. & Bork, P. SAM as a protein interaction domain involved in developmental regulation *Protein Sci* 6, 249-53 (1997).

2. Jousset, C., *et al.* A domain of TEL conserved in a subset of ETS proteins defines a specific oligomerization interface essential to the mitogenic properties of the TEL-PDGFR beta oncoprotein *Embo J* 16, 69-82 (1997).

3. Peterson, A.J., *et al.* A domain shared by the Polycomb group proteins Scm and ph mediates heterotypic and homotypic interactions *Mol Cell Biol* 17, 6683-92 (1997).

4. Tu, H., Barr, M., Dong, D.L. & Wigler, M. Multiple regulatory domains on the Byr2 protein kinase *Mol Cell Biol* 17, 5876-87 (1997).

5. Kyba, M. & Brock, H.W. The SAM domain of polyhomeotic, RAE28, and scm mediates specific interactions through conserved residues *Dev Genet* 22, 74-84 (1998).

6. Golub, T.R., Barker, G.F., Lovett, M. & Gilliland, D.G. Fusion of PDGF receptor beta to a novel ets-like gene, tel, in chronic myelomonocytic leukemia with t(5;12) chromosomal translocation *Cell* 77, 307-16 (1994).

7. Golub, T.R., *et al.* Oligomerization of the ABL tyrosine kinase by the Ets protein TEL in human leukemia *Mol Cell Biol* 16, 4107-16 (1996).

8. Lacronique, V., *et al.* A TEL-JAK2 fusion protein with constitutive kinase activity in human leukemia *Science* 278, 1309-12 (1997).

9. Golub, T.R., *et al.* Fusion of the TEL gene on 12p13 to the AML1 gene on 21q22 in acute lymphoblastic leukemia *Proc Natl Acad Sci U S A* 92, 4917-21 (1995).

10. Henkemeyer, M., *et al.* Nuk controls pathfinding of commissural axons in the mammalian central nervous system *Cell* 86, 35-46 (1996).

11. Orioli, D., Henkemeyer, M., Lemke, G., Klein, R. & Pawson, T. Sek4 and Nuk receptors cooperate in guidance of commissural axons and in palate formation *Embo J* 15, 6035-49 (1996).

12. Krull, C.E., *et al.* Interactions of Eph-related receptors and ligands confer rostrocaudal pattern to trunk neural crest migration *Curr Biol* 7, 571-80 (1997).

13. Xu, Q., Alldus, G., Macdonald, R., Wilkinson, D.G. & Holder, N. Function of the Eph-related kinase rtk1 in patterning of the zebrafish forebrain *Nature* 381, 319-22 (1996).

14. Wang, H.U., Chen, Z.F. & Anderson, D.J. Molecular distinction and angiogenic interaction between embryonic arteries and veins revealed by ephrin-B2 and its receptor Eph-B4 *Cell* 93, 741-53 (1998).

15. Stein, E., *et al.* Eph receptors discriminate specific ligand oligomers to determine alternative signaling complexes, attachment, and assembly responses *Genes Dev* 12, 667-78 (1998).

16. Lackmann, M., *et al.* Ligand for EPH-related kinase (LERK) 7 is the preferred high affinity ligand for the HEK receptor *J Biol Chem* 272, 16521-30 (1997).

17. Davis, S., *et al.* Ligands for EPH-related receptor tyrosine kinases that require membrane attachment or clustering for activity *Science* 266, 816-9 (1994).

18. Lackmann, M., *et al.* Distinct subdomains of the EphA3 receptor mediate ligand binding and receptor dimerization *J Biol Chem* 273, 20228-37 (1998).

19. Hock, B., *et al.* PDZ-domain-mediated interaction of the eph-related receptor tyrosine kinase EphB3 and the ras-binding protein AF6 depends on the kinase activity of the receptor *Proc Natl Acad Sci U S A* 95, 9779-84 (1998).

20. Ponting, C.P. SAM: a novel motif in yeast sterile and Drosophila polyhomeotic proteins *Protein Sci* 4, 1928-30 (1995).

21. Klambt, C. The Drosophila gene pointed encodes two ETS-like proteins which are involved in the development of the midline glial cells *Development* 117, 163-76 (1993).

22. Serra-Pages, C., Medley, Q.G., Tang, M., Hart, A. & Streuli, M. Liprins, a family of LAR transmembrane protein-tyrosine phosphatase- interacting proteins *J Biol Chem* 273, 15611-20 (1998).

23. Kuriyan, J. & Cowburn, D. Modular Peptide Binding Domains *Annu. Rev. Biophys. Biomol. Struct.* 26, 259-288 (1997).

24. Stein, E., Cerretti, D.P. & Daniel, T.O. Ligand activation of ELK receptor tyrosine kinase promotes its association with Grb10 and Grb2 in vascular endothelial cells *J Biol Chem* 271, 23588-93 (1996).

25. Terwilliger, T.C., Kim, S.H. & Eisenberg, D. Generalized method of determining heavy-atom positions using the difference Patterson function *Acta Crytallogr. Sect. A* 43, 1-5 (1987).

26. Furey, W. & Swaminathan, S. PHASES *in American Crystallographic Association Meeting Abstracts* 73 (1990).

27. Jones, T.A., Zou, J.Y., Cowan, S.W. & Kjeldgaard, M. Improved methods for building protein models in electron density maps and the location of errors in these models *Acta Crystallogr.* A47, 110-119 (1991).

28. Brünger, A:T. (Yale University, New Haven, CT, 1992).

29. Carson, M. Ribbons 2.0 *Journal of applied Crystallography* 24, 958 (1991).

30. Nicholls, A., Sharp, K.A. & Honig, B. Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons *Proteins: Struct. Funct. and Genetics* 11, 281-296 (1991).

Claims

WE CLAIM:

1. A purified three dimensional structure of a polypeptide corresponding to one or more SAM domains.
2. A three dimensional structure as claimed in claim 1, wherein the SAM domain is a SAM domain of an Eph receptor.
3. A three dimensional structure as claimed in claim 2 wherein the Eph receptor is EphA.
4. A three dimensional structure as claimed in claim 1 complexed with one or more compounds.
5. A three dimensional structure as claimed in claim 1 comprising one or more heavy metal atoms.
6. A purified crystalline form of a polypeptide corresponding to one or more SAM domains.
7. A crystalline form as claimed in claim 6 having dimensions of about a=b= 77.14 ± .03 angstroms, c= 24.3 ± .04 angstroms.
8. A crystalline form as claimed in claim 7 having the co-ordinates set out in Table 2.
9. A method of forming a crystalline form as claimed in claim 6 comprising
   (a) mixing a volume of a SAM domain with a reservoir solution; and
   (b) incubating the mixture obtained in step (a) over the reservoir solution in a closed container under conditions suitable for crystallization.
10. A method of determining three dimensional structures of polypeptides with SAM domains of unknown structure comprising the step of applying the structural atomic coordinates of a three dimensional structure as claimed in claim 1 or a crystalline form as claimed in claim 7 or 8.
11. A method for identifying a potential modulator of a SAM domain of an Eph receptor function comprising docking a computer representation of a structure of a compound with a computer representation of a structure of one or more SAM domains of an Eph receptor that is defined by the atomic structural coordinates of the three dimensional structure as claimed in claim 2 or a crystalline form as claimed in claim 7 or 8.
12. A method as claimed in claim 11 comprising the following steps:
   (a) docking a computer representation of a compound from a computer data base with a computer representation of a selected site on a three dimensional structure of a SAM domain of an Eph receptor as claimed in claim 2 or a crystalline form as claimed in claim 7 or 8 to obtain a complex;
   (b) determining a conformation of the complex with a favourable geometric fit and favourable complementary interactions; and
   (c) identifying compounds that best fit the selected site as potential modulators of SAM domain function.
13. A method as claimed in claim 11, comprising the following steps:
   (a) modifying a computer representation of a compound complexed with a selected site on a three dimensional structure of a SAM domain of an Eph receptor as claimed in claim 2 or a crystalline form as claimed in claim 7 or 8, by deleting or adding a chemical group or groups;

(b) determining a conformation of the complex with a favourable geometric fit and favourable complementary interactions; and

(c) identifying a compound that best fits the selected site as a potential modulator of a SAM domain.

14. A method as claimed in claim 11 comprising the following steps:

(a) selecting a computer representation of a compound complexed with a selected site on a three dimensional structure of a SAM domain of an Eph receptor as claimed in claim 2 or a crystalline form as claimed in claim 7 or 8; and

(b) searching for molecules in a data base that are similar to the compound using a searching computer program, or replacing portions of the compound with similar chemical structures from a data base using a compound building computer program.

15. A potential modulator of a function of a SAM domain of an Eph receptor identified by a method as claimed in any one of claims 11 to 14.

16. A method of treating a disease associated with a SAM domain of an Eph receptor with inappropriate activity in a cellular organism, comprising:

(a) administering a crystalline form of a polypeptide as claimed in claim 6 or a modulator identified using a method as claimed in any one of claims 11 to 14, in an acceptable pharmaceutical preparation; and

(b) activating or inhibiting a SAM domain function to treat the disease.

17. A method as claimed in claim 16 wherein the disease is a cell proliferative disease or disease associated with the nervous system.

18. A peptide of the formula I which mediates SAM domain function:

$$X\text{-}X^1\text{-}X^2\text{-}X^3\text{-}X^4\text{-}X^5\text{-}X^6 \qquad I$$

wherein X and X^6 represent 0 to 70, preferably 0 to 50 amino acids, more preferably 2 to 20 amino acids, and X^1 represents Leu, Phe, Asp, Ala, Glu, or Gly, preferably Leu or Gly, X^2 represents Glu, Asp, Ser, Ile, Ala, Arg, Lys, and Gln, preferably Glu or Asp, X^3 represents Ala, Val, Glu, Phe, Ser, Ile, Met, Leu, His, Gln, Arg, or Asp preferably Ala, Val, or Phe, X^4 is Val, Leu, Met, Phe, and Ile, preferably Val or Leu, or Phe, X^5 is Val, Ser, Leu, Asp, Ala, Pro, Asn, Lys, or Cys, preferably Val or Ser.

19. A peptide as claimed in claim 18 wherein X represents TT, ID, TS, DD, GYTT (SEQ ID. NO. 38), AAGYTT (SEQ ID. NO. 39), FTAAGYTT (SEQ ID. NO. 40), DNFTAAGYTT (SEQ ID. NO. 41), or YKDNFTAAGYTT (SEQ ID. NO. 42).

20. A peptide as claimed in claim 18 wherein X^6 represents HM, HMSQ (SEQ ID. NO. 43), HMSQD (SEQ ID. NO. 44), HMSQDD (SEQ ID. NO. 45), HMSQDDLA (SEQ ID. NO. 46), QMMM (SEQ ID. NO. 47), QMMMED (SEQ ID. NO. 48), QMMMEDLL (SEQ ID. NO. 49), DITE (SEQ ID. NO. 50), DITEED (SEQ ID. NO. 51), DITEEDL (SEQ ID. NO. 52), NLTE (SEQ ID. NO. 53), NLTEND (SEQ ID. NO. 54), or NLTENDI (SEQ ID. NO. 55).

21. A peptide of the formula I as claimed in claim 18 which is LEAVV (SEQ ID. NO. 56), TTLEAVV (SEQ ID. NO. 57), LEAVVHM (SEQ ID. NO. 58), LEAVVHMSQ (SEQ ID. NO. 59), LEAVVHMSQD (SEQ ID. NO. 60), LEAVVHMSQDDL (SEQ ID. NO. 61), LEAVVHMSQDDLAR (SEQ ID. NO. 62), TTLEAVVHMS (SEQ ID. NO. 63), TTLEAVVHMSQD (SEQ ID. NO. 64), TTLEAVVHMSQDDL (SEQ ID. NO. 65), TTLEAVVHMSQDDLAR (SEQ ID. NO. 66), GYTTLEAVV (SEQ ID. NO. 67), GYTTLEAVVHMS (SEQ ID. NO. 68), GYTTLEAVVHMSQD (SEQ ID. NO. 69), GYTTLEAVVHMSQDDL (SEQ ID. NO. 70), GYTTLEAVVHMSQDDLAR (SEQ ID. NO. 71), FDVVS (SEQ ID. NO. 72), FDVVSQ (SEQ ID. NO. 73), FDVVSQMM (SEQ ID. NO. 74), FDVVSQMMME (SEQ ID. NO. 75), FDVVSQMMMEDIL (SEQ ID. NO. 76), TSFDVVS (SEQ ID. NO. 77), TSFDVVSQ (SEQ ID. NO. 78), TSFDVVSQMM (SEQ ID. NO. 79), TSFDVVSQMMME (SEQ ID. NO. 80), TSFDVVSQMMMEDIL (SEQ ID. NO. 81), LEFLS (SEQ ID. NO. 82), LEFLSD (SEQ ID. NO. 83), LEFLSDIT (SEQ ID. NO. 84), LEFLSDITEE (SEQ ID. NO. 85), LEFLSDITEEDL (SEQ ID. NO. 86), DDLEFLS (SEQ ID. NO. 87), GWDDLEFLS (SEQ ID. NO. 88), DDLEFLSD (SEQ ID. NO. 89), DDLEFLSDIT (SEQ ID. NO. 90), DDLEFLSDITEE (SEQ ID. NO. 91), DDLEFLSDITEEDL (SEQ ID. NO. 92), GARFL (SEQ ID. NO. 93), GARFLN (SEQ ID. NO. 94), GARFLNLT (SEQ ID. NO. 95), GARFLNLTEN (SEQ ID. NO. 96), and IDGARFL (SEQ ID. NO. 97).

22. A peptide of the formula II which mediates SAM domain function:

$$X^7\text{-}X^8\text{-}X^9\text{-}X^{10}\text{-}X^{11}\text{-}X^{12}\text{-}X^{13}\text{-}X^{14}\text{-}X^{15}\text{-}X^{16} \qquad \text{II}$$

wherein X^7 and X^{16} represent 0 to 70, preferably 0 to 50 amino acids, more preferably 2 to 20 amino acids, and X^8 represents Met, Ile, Ser, Leu, Asn, Phe, or Val, preferably Met, X^9 represents Arg, Ser, Lys, Met, Leu, Glu, Gln, or Asn, preferably Gln or Arg, X^{10} represents Thr, Ala, Arg, Leu, Ser, Glu, Asp, Met, Lys, Gln, or Gly, preferably Thr, Ala, or Glu, X^{11} represents Gln, Ser, Glu, Leu, Phe, Asp, Thr, Arg, preferably Gln or Arg, X^{12} represents Met, Ala, Ile, Asn, Ser, Arg, Thr, Pro, Leu, Gln, Val, Lys, preferably Met or Arg, X^{13} represents Gln, Asn, Pro, Ser, Tyr, Glu, Leu, Arg, or Lys, preferably Gln, Asn, or Arg, X^{14} represents Gln, Ala, Pro, Asp, Leu, Lys, Ile, Glu, Arg, or Asn, preferably Gln or Ile, and X^{15} represents Met, Ile, Val, His, Ser, Arg, Lys, Phe, Cys, Glu, Tyr, Ala, Ile, Trp, or Leu.

23. A peptide of the formula II as claimed in claim 22 wherein X^7 represents QA, QV, NK, SVQA (SEQ ID. NO. 98), LSSVQA (SEQ ID. NO. 99), ILSSVQA (SEQ ID. NO. 100), NKILSSVQA (SEQ ID. NO. 101), HQNKILSSVQA (SEQ ID. NO. 102), THQNKILSSVQA (SEQ ID. NO. 103), ENIK (SEQ ID. NO. 104), SQEINK (SEQ ID. NO. 105), KLSQEINK (SEQ ID. NO. 106), ILNSIQV (SEQ ID. NO. 107), or NSIQV (SEQ ID. NO. 108).

24. A peptide of the formula II as claimed in claim 22 wherein X^{16} is HG, QS, HGRM (SEQ ID. NO. 109), HGRMVP (SEQ ID. NO. 110), QSVEV (SEQ ID. NO. 111), or TRKP (SEQ ID. NO. 112).

25. A peptide of the formula II as claimed in claim 22 which is MRTQMQQM (SEQ ID. NO. 113), QAMRTQMQQM (SEQ ID. NO. 114), SVQAMRTQMQQM (SEQ ID. NO. 115), LSSVQAMRTQMQQM (SEQ ID. NO. 116), ILSSVQAMRTQMQQM (SEQ ID. NO. 117), MRTQMQQMHG (SEQ ID. NO. 118), MRTQMQQMHGRM (SEQ ID. NO. 119), MRTQMQQMHGRMVPV (SEQ ID. NO. 120), NEERRSIF (SEQ ID. NO. 121), INKNEERRSIF (SEQ ID. NO. 122), NEERRSIFTRKP (SEQ ID. NO. 123). MRAQMNQI (SEQ ID. NO. 124), MRAQMNQIQS (SEQ ID. NO. 125), MRAQMNQIQSVEV (SEQ ID. NO. 126).

26. A peptide which mediates SAM domain function comprising VVSV (SEQ ID. NO. 21), SAVVSV (SEQ ID. NO.22), FSAVV (SEQ ID. NO.23), FSAVVSV (SEQ ID. NO. 24), FSAVVSVGD (SEQ ID. NO. 25), VVSVGDWL (SEQ ID. NO. 26), FNTV (SEQ ID. NO. 27), FNTVDE (SEQ ID. NO. 28), FNTVDEWL (SEQ ID. NO. 29), TSFNTVDEWL (SEQ ID. NO. 30), TSFNTV (SEQ ID. NO. 31), YTSFNTV (SEQ ID. NO. 32), RSEV (SEQ ID. NO. 33), RSEVLG (SEQ ID. NO. 34), RSEVLGVD (SEQ ID. NO. 35), VPFRSEV (SEQ ID. NO. 36), and VPFRSEVLGW (SEQ ID. NO. 37).

27. A pharmaceutical composition comprising a peptide as claimed in any one of claims 18 to 26 and a pharmaceutically acceptable carrier, diluent or excipient.

FIGURE 1A

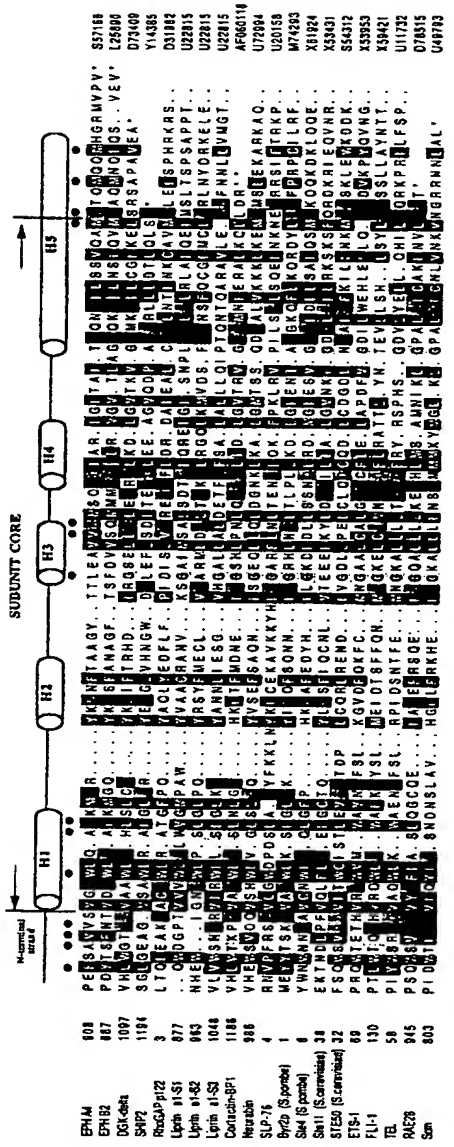

SUBUNIT CORE
H1
H2
H3
H4
H5
N-terminal extent
EPHA4
EPHB2
DOK-like
SHP2
RasGAP-p120
Lipin α1-S1
Lipin α1-S2
Lipin α1-S3
Centaurin-β1
Neurabin
SLP-76
Byr2p (S.pombe)
Ste4 (S.pombe)
Sam1 (S.cerevisiae)
STE50 (S.cerevisiae)
ETS-1
FLI-1
TEL
RAE28
Sam

FIGURE 1B

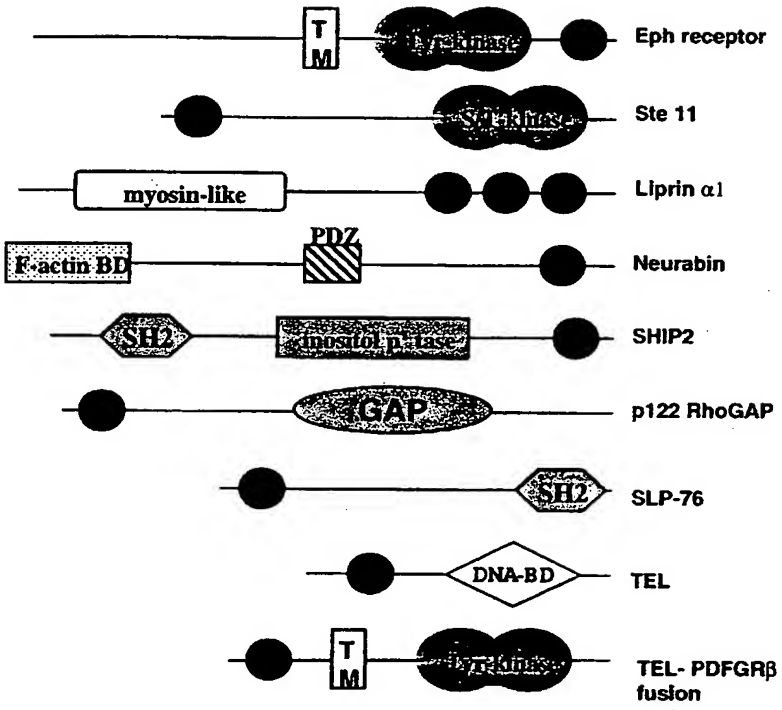

T
M
Eph receptor
Ste 11
myosin-like
Liprin α1
PDZ
F-actin BD
Neurabin
SH2
SHIP2
GAP
p122 RhoGAP
SH2
SLP-76
DNA-BD
TEL
T
M
TEL- PDFGRβ
fusion

FIGURE 2A

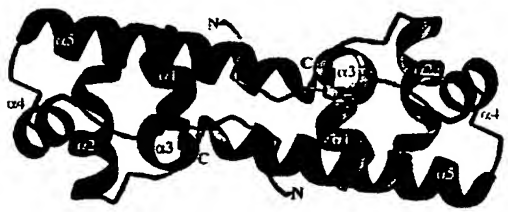


FIGURE 2B

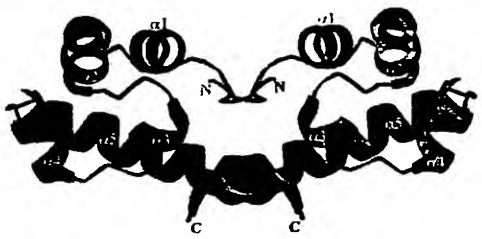


FIGURE 2C

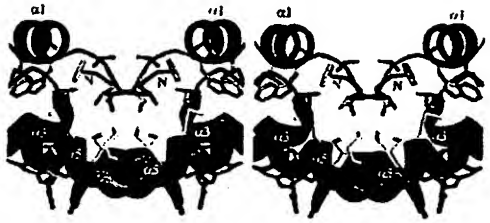


FIGURE 3A

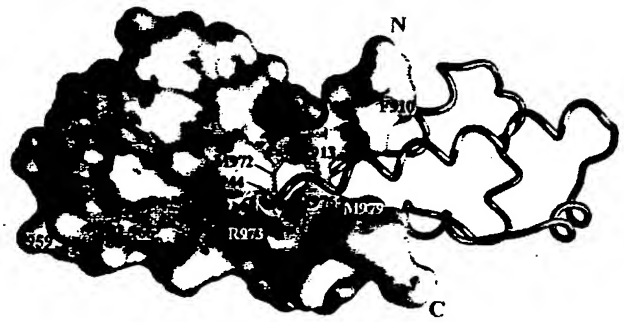
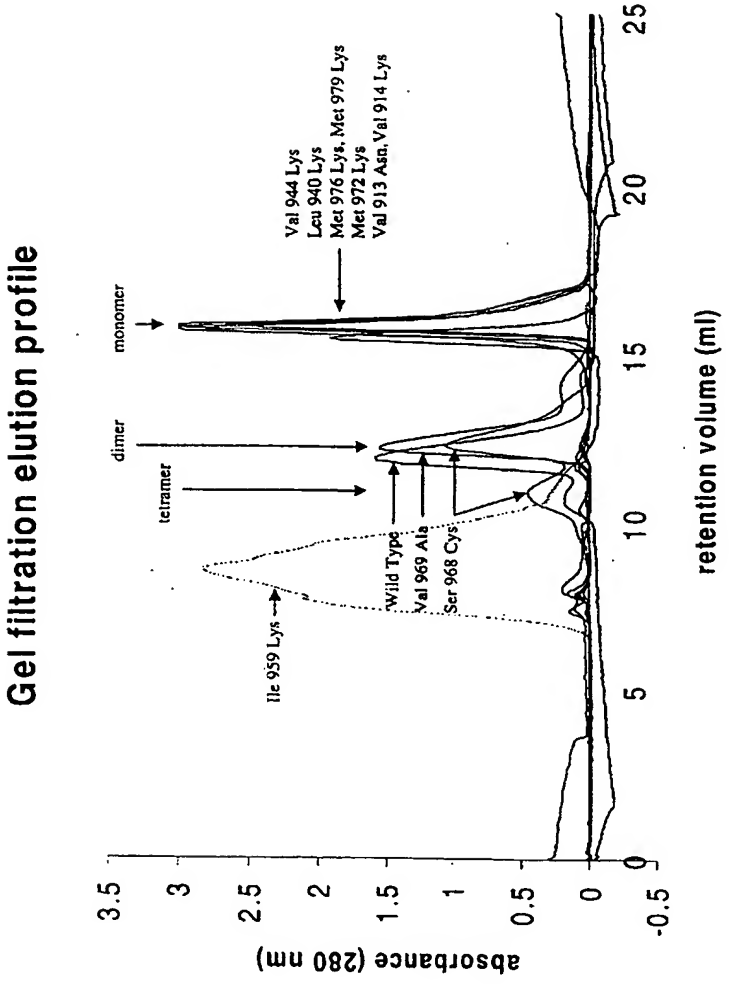


FIGURE 3B

FIGURE 4


Gel filtration elution profile
monomer
dimer
tetramer
Val 944 Lys
Leu 940 Lys
Met 976 Lys, Met 979 Lys
Met 972 Lys
Val 913 Asn, Val 914 Lys
Wild Type
Val 969 Ala
Ser 968 Cys
Ile 959 Lys
absorbance (280 nm)
retention volume (ml)
3.5
3
2.5
2
1.5
1
0.5
0
-0.5
0
5
10
15
20
25

Sequence Listing

SEQ. ID. NO. 1

EPH A4

PEFSAVVSVGDWLQAIKMDRYKDNFTAAGYTTLEAVVHMSQDDLARIGITAITHQNKILSSVQ
AMRTQMQQMHGRMVPV

SEQ. ID. NO. 2

EPH B2

PDYTSFNTVDEWLEAIKMGQYKESFANAGFTSFDVVSQMMMEDILRVGVTLAGHQKKILNSI
QVMRAQMNQIQSVEV

SEQ. ID. NO. 3

DGK-delta

VHLVGTEEVAAWLEHLSLCEYKDIFTRHDIRGSELLHLERRDLKDLGVTKVGHMKRILCGIKE
LSRSAPAVEA

SEQ. ID. NO. 4

SHIP2

SGLGEAGMSAWLRAIGLERYEEGLVHNGWDDLEFLSDITEEDLEEAGVQDPAHKRLLLDTLQ
LSK

SEQ. ID. NO. 5

RhoGAP p122

LTQIEAKEACDWLRATGFPQYAQLYEDFLFPIDISLVKREHDFLDRDAIEALCRRLNTLNKCAV
MKLEISPHRKRS

SEQ. ID. NO. 6

Liprin a1-S1

QWDGPTVVVWLELWVGMPAWYVAACRANVKSGAIMSALSDTEIQREIGISNPLHRLKLRLAI
QEIMSLTSPSAPPT

SEQ. ID. NO. 7

Liprin a1-S2

NIIEWIGNEWLPSLGLPQYRSYFMECLVDARMLDHLTKKDLRGQLKMVDSFHRNSFQCGIMC
LRRLNYDRKELE

SEQ. ID. NO. 8

Liprin a1-S3

VLVWSNDRVIRWILSIGLKYANNLIESGVHGALLALDETFDFSALALLLQIPTQNTQARAVLER
EFNNLLVMGT

SEQ. ID. NO. 9

Cortactin-BP1

VHLWTKPDVADWLESLNLGEHKETFMDNEIDGSHLPNLQKEDLIDLGVTRVGHRMNIERALK
QLLDR

SEQ. ID. NO. 10

Neurabin

VHEWSVQQVSHWLVGLSLDQYVSEFSAQNISGEQLLQLDGNKLKALGMTSSQDRALVKKKL
KEMKMSLEKARKAQ

SEQ. ID. NO. 11

SLP-76

RNVPFRSEVLGWDPDSLADYFKKLNYDCEKAVKKYHIDGARFLNLTENDIQKFPKLRVPILSK
LSQEINKNEERRSIFTRKP

SEQ. ID. NO. 12

Byr2p (S.pombe)

MEYYTSKEVAEWLKSIGLEKYIEQFSQNNIEGRHLNHLTLPLLKDLGIENTAKGKQFLKQDYL
REFPRPCILRF

SEQ. ID. NO. 13

Ste4(S.pombe)

YWNWNNEAVCNWIEQLGFPHKEAFEDYHILGKDIDLLSSNDLRDMGIESVGHRIDILSAIQSM
KKQKDKLQQE

SEQ. ID. NO. 14

Ste11 (S.cerevisiae)

EKTNDLPFVQLFLEEIGCTQYLDSFIQCNLVTEEEIKYLDKDILIALGVNKIGDRLKILRKSKSFQ
RDKRIEQVNR

SEQ. ID. NO. 15

STE50 (S. cerevisiae)

FSQWSVDDVITWCISTLEVEETDPLCQRLRENDIVGDLLPELCLQDLCDGDLNKAIKFKILINK
MRDSKLEWKDDK

SEQ. ID. NO. 16

ETS-1

PRQWTETHVRDWVMWAVNEFSLKGVDFQKFCMNGAALCALGKDCFLELAPDFVGDILWEH
LEILQKEDVKPYQVNG

SEQ. ID. NO. 17

FLI-1

PTLWTQEHVRQWLEWAIKEYSLMEIDTSFFQNMDGKELCKMNKEDFLRATTLYNTEVLLSHL
SYLRESSLLAYNTT

SEQ. ID. NO. 18

TEL

PIYWSRDDVAQWLKWAENEFSLRPIDSNTFEMNGKALLLLTKEDFRYRSPHSGDVLYELLQHI
LKQRKPRILFSP

SEQ. ID. NO. 19

RAE28

PSQWSVEEVYEFIASLQGCQEIAEEFRSQEIDGQALLLLKEEHLMSAMNIKLGPALKICAKINVL
KET

SEQ. ID. NO. 20

Scm

PIDWTIEEVIQYIESNDNSLAVHGDLFRKHEIDGKALLLLNSEMMMKYMGLKLGPALKICNLV
NKVNGRRNNLAL

SEQ. ID. NO. 21

VVSV

SEQ ID. NO. 22

SAVVSV

SEQ ID. NO.23

FSAVV

SEQ ID. NO.24

FSAVVSV

SEQ ID. NO. 25

FSAVVSVGD

SEQ ID. NO. 26

VVSVGDWL

SEQ ID. NO. 27

FNTV

SEQ ID. NO. 28

FNTVDE

SEQ ID. NO. 29

FNTVDEWL

SEQ ID. NO. 30

TSFNTVDEWL

SEQ ID. NO. 31

TSFNTV

SEQ ID. NO. 32

YTSFNTV

SEQ ID. NO. 33

RSEV

SEQ ID. NO. 34

RSEVLG

SEQ ID. NO. 35

RSEVLGWD

SEQ ID. NO. 36

VPFRSEV

SEQ ID. NO. 37

VPFRSEVLGW

SEQ ID. NO. 38

GYTT

SEQ ID. NO. 39

AAGYTT

SEQ ID. NO. 40

FTAAGYTT

SEQ ID. NO. 41

DNFTAAGYTT

SEQ ID. NO. 42

YKDNFTAAGYTT

SEQ ID. NO. 43

HMSQ

SEQ ID. NO. 44

HMSQD

SEQ ID. NO. 45

HMSQDD

SEQ ID. NO. 46

HMSQDDLA

SEQ ID. NO. 47

QMMM
SEQ ID. NO. 48
QMMMED
SEQ ID. NO. 49
QMMMEDLL
SEQ ID. NO. 50
DITE
SEQ ID. NO. 51
DITEED
SEQ ID. NO. 52
DITEEDL
SEQ ID. NO. 53
NLTE
SEQ ID. NO. 54
NLTEND
SEQ ID. NO. 55
NLTENDI
SEQ ID. NO. 56
LEAVV
SEQ ID. NO. 57
TTLEAVV
SEQ ID. NO. 58
LEAVVHM
SEQ ID. NO. 59
LEAVVHMSQ
SEQ ID. NO. 60
LEAVVHMSQD
SEQ ID. NO. 61
LEAVVHMSQDDL
SEQ ID. NO. 62
LEAVVHMSQDDLAR
SEQ ID. NO. 63
TTLEAVVHMS
SEQ ID. NO. 64
TTLEAVVHMSQD
SEQ ID. NO. 65
TTLEAVVHMSQDDL
SEQ ID. NO. 66
TTLEAVVHMSQDDLAR
SEQ ID. NO. 67

GYTTLEAVV
SEQ ID. NO. 68
GYTTLEAVVHMS
SEQ ID. NO. 69
GYTTLEAVVHMSQD
SEQ ID. NO. 70
GYTTLEAVVHMSQDDL
SEQ ID. NO. 71
GYTTLEAVVHMSQDDLAR
SEQ ID. NO. 72
FDVVS
SEQ ID. NO. 73
FDVVSQ
SEQ ID. NO. 74
FDVVSQMM
SEQ ID. NO. 75
FDVVSQMMME
SEQ ID. NO. 76
FDVVSQMMMEDIL
SEQ ID. NO. 77
TSFDVVS
SEQ ID. NO. 78
TSFDVVSQ
SEQ ID. NO. 79
TSFDVVSQMM
SEQ ID. NO. 80
TSFDVVSQMMME
SEQ ID. NO. 81
TSFDVVSQMMMEDIL
SEQ ID. NO. 82
LEFLS
SEQ ID. NO. 83
LEFLSD
SEQ ID. NO. 84
LEFLSDIT
SEQ ID. NO. 85
LEFLSDITEE
SEQ ID. NO. 86
LEFLSDITEEDL
SEQ ID. NO. 87

DDLEFLS
SEQ ID. NO. 88
GWDDLEFLS
SEQ ID. NO. 89
DDLEFLSD
SEQ ID. NO. 90
DDLEFLSDIT
SEQ ID. NO. 91
DDLEFLSDITEE
SEQ ID. NO. 92
DDLEFLSDITEEDL
SEQ ID. NO. 93
GARFL
SEQ ID. NO. 94
GARFLN
SEQ ID. NO. 95
GARFLNLT
SEQ ID. NO. 96
GARFLNLTEN
SEQ ID. NO. 97
IDGARFL
SEQ ID. NO. 98
SVQA
SEQ ID. NO. 99
LSSVQA
SEQ ID. NO. 100
ILSSVQA
SEQ ID. NO. 101
NKILSSVQA
SEQ ID. NO. 102
HQNKILSSVQA
SEQ ID. NO. 103
THQNKILSSVQA
SEQ ID. NO. 104
ENIK
SEQ ID. NO. 105
SQEINK
SEQ ID. NO. 106
KLSQEINK
SEQ ID. NO. 107

ILNSIQV
SEQ ID. NO. 108
NSIQV
SEQ ID. NO. 109
HGRM
SEQ ID. NO. 110
HGRMVP
SEQ ID. NO. 111
QSVEV
SEQ ID. NO. 112
TRKP
SEQ ID. NO. 113
MRTQMQQM
SEQ ID. NO. 114
QAMRTQMQQM
SEQ ID. NO. 115
SVQAMRTQMQQM
SEQ ID. NO. 116
LSSVQAMRTQMQQM
SEQ ID. NO. 117
ILSSVQAMRTQMQQM
SEQ ID. NO. 118
MRTQMQQMHG
SEQ ID. NO. 119
MRTQMQQMHGRM
SEQ ID. NO. 120
MRTQMQQMHGRMVPV
SEQ ID. NO. 121
NEERRSIF
SEQ ID. NO. 122
INKNEERRSIF
SEQ ID. NO. 123
NEERRSIFTRKP
SEQ ID. NO. 124
MRAQMNQI
SEQ ID. NO. 125
MRAQMNQIQS
SEQ ID. NO. 126
MRAQMNQIQSVEV

INTERNATIONAL SEARCH REPORT

Inte ...nal Application No
PCT/CA 99/01209

A. CLASSIFICATION OF SUBJECT MATTER
IPC 7 C07K14/705 C07K14/47 G06F17/50 A61K38/17 A61P25/00

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 7 C07K G06F A61K A61P

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
P,X	STAPLETON DAVID ET AL: "The crystal structure of an Eph receptor SAM domain reveals a mechanism for modular dimerization." NATURE STRUCTURAL BIOLOGY JAN., 1999, vol. 6, no. 1, January 1999 (1999-01), pages 44-49, XP002134936 ISSN: 1072-8368 the whole document	1-16

-/—

[X] Further documents are listed in the continuation of box C.

[X] Patent family members are listed in annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
6 April 2000	25/04/2000

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Cervigni, S

Form PCT/ISA/210 (second sheet) (July 1992)

INTERNATIONAL SEARCH REPORT

Inter nal Application No
PCT/CA 99/01209

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication,where appropriate, of the relevant passages	Relevant to claim No.
P,X	THANOS CHRISTOPHER D ET AL: "Oligomeric structure of the human EphB2 receptor SAM domain." SCIENCE (WASHINGTON D C) FEB. 5, 1999, vol. 283, no. 5403, 5 February 1999 (1999-02-05), pages 833-836, XP002134937 ISSN: 0036-8075 the whole document	1-16
A	HIMANEN JUHA-PEKKA ET AL: "Crystal structure of the ligand-binding domain of the receptor tyrosine kinase EphB2." NATURE (LONDON) DEC. 3, 1998, vol. 396, no. 6710, 3 December 1998 (1998-12-03), pages 486-491, XP002134939 ISSN: 0028-0836	
A	SCHULTZ JOERG ET AL: "SAM as a protein interaction domain involved in developmental regulation." PROTEIN SCIENCE 1997, vol. 6, no. 1, 1997, pages 249-253, XP002134938 ISSN: 0961-8368	
A	WO 98 47088 A (BLOM NICO ;UNIV MONTREAL (CA); SYGUSCH JURGEN (CA)) 22 October 1998 (1998-10-22)	
A	LYBRAND T P: "LIGAND-PROTEIN DOCKING AND RATIONAL DRUG DESIGN" CURRENT OPINION IN STRUCTURAL BIOLOGY,GB,CURRENT BIOLOGY LTD., LONDON, vol. 5, 1 January 1995 (1995-01-01), pages 224-228, XP000764926 ISSN: 0959-440X	
A	WO 95 27060 A (REGENERON PHARMA) 12 October 1995 (1995-10-12)	
A	WO 97 14966 A (MOUNT SINAI HOSPITAL CORP) 24 April 1997 (1997-04-24)	
X	WO 96 40189 A (GLAXO GROUP LTD ;DOWER WILLIAM J (US); BARRETT RONALD W (US); CWIR) 19 December 1996 (1996-12-19) page 29, line 18	18
X	WO 93 10816 A (UNIV TEXAS) 10 June 1993 (1993-06-10) D39 page 13; table 2	18

1

Form PCT/ISA/210 (continuation of second sheet) (July 1992)

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CA 99/01209

Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This International Search Report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. [X] Claims Nos.: 1-5
because they relate to subject matter not required to be searched by this Authority, namely:

see FURTHER INFORMATION sheet PCT/ISA/210

2. [X] Claims Nos.: 18-20,22-24,26
because they relate to parts of the International Application that do not comply with the prescribed requirements to such an extent that no meaningful International Search can be carried out, specifically:

see FURTHER INFORMATION sheet PCT/ISA/210

3. [ ] Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

1. [ ] As all required additional search fees were timely paid by the applicant, this International Search Report covers all searchable claims.

2. [ ] As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.

3. [ ] As only some of the required additional search fees were timely paid by the applicant, this International Search Report covers only those claims for which fees were paid, specifically claims Nos.:

4. [ ] No required additional search fees were timely paid by the applicant. Consequently, this International Search Report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

[ ] The additional search fees were accompanied by the applicant's protest.

[ ] No protest accompanied the payment of additional search fees.

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

Continuation of Box I.1

Although claims 16-17 are directed to a method of treatment of the human/animal body, the search has been carried out and based on the alleged effects of the compound/composition.

Continuation of Box I.1

Claims Nos.: 1-5

Rule 39.1(i) PCT - scientific and mathematical theories

Continuation of Box I.2

Claims Nos.: 18-20,22-24,26

Present claims 18-20,22-24 and 26 relate to an extremely large number of possible compounds. Support within the meaning of Article 6 PCT and/or disclosure within the meaning of Article 5 PCT is to be found, however, for only a very small proportion of the compounds claimed. In the present case, the claims so lack support, and the application so lacks disclosure, that a meaningful search over the whole of the claimed scope is impossible. Consequently, the search has been carried out for those parts of the claims which appear to be supported and disclosed, namely those parts relating to the compounds listed claims in 21 and 25.

The applicant's attention is drawn to the fact that claims, or parts of claims, relating to inventions in respect of which no international search report has been established need not be the subject of an international preliminary examination (Rule 66.1(e) PCT). The applicant is advised that the EPO policy when acting as an International Preliminary Examining Authority is normally not to carry out a preliminary examination on matter which has not been searched. This is the case irrespective of whether or not the claims are amended following receipt of the search report or during any Chapter II procedure.

INTERNATIONAL SEARCH REPORT

Information on patent family members

Inter nal Application No
PCT/CA 99/01209

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 9847088 A	22-10-1998	US 5866343 A	02-02-1999
		AU 6915198 A	11-11-1998
WO 9527060 A	12-10-1995	US 5747033 A	05-05-1998
		AU 691915 B	28-05-1998
		AU 2278995 A	23-10-1995
		CA 2187167 A	12-10-1995
		EP 0758381 A	19-02-1997
		JP 9511401 T	18-11-1997
		ZA 9502762 A	20-02-1996
WO 9714966 A	24-04-1997	JP 11515105 T	21-12-1999
WO 9640189 A	19-12-1996	AU 6046696 A	30-12-1996
		HR 960256 A	28-02-1998
		AU 6381896 A	30-12-1996
		EP 0842293 A	20-05-1998
		WO 9640987 A	19-12-1996
		AU 704215 B	15-04-1999
		AU 6163496 A	30-12-1996
		BR 9608587 A	05-01-1999
		CA 2223449 A	19-12-1996
		CN 1192749 A	09-09-1998
		CZ 9703897 A	17-06-1998
		EP 0885242 A	23-12-1998
		HU 9900921 A	28-07-1999
		JP 10507776 T	28-07-1998
		NO 975705 A	05-02-1998
		NZ 310778 A	28-10-1999
		PL 323917 A	27-04-1998
		WO 9640750 A	19-12-1996
		US 5869451 A	09-02-1999
WO 9310816 A	10-06-1993	AT 190226 T	15-03-2000
		AU 666160 B	01-02-1996
		AU 3233993 A	28-06-1993
		CA 2124691 A	10-06-1993
		EP 0671947 A	20-09-1995
		EP 0968721 A	05-01-2000
		JP 7502729 T	23-03-1995